INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours Full marks : 100

(a) Figures in the margin indicate full marks.

GROUP A

(Answer any three questions from this group)

1. i) Enumerate the different sources providing data on the production of foodgrains in India, and mention the agencies responsible for their publication.

ii) How would you compute per capita availability of foodgrains in India? Indicate clearly the information needed for this purpose.

iii) Is it possible to check your estimate in (ii) against survey data on per capita consumption of foodgrains? (6+7+3)=16

2. Select one important publication relating to the work of any two of the following institutions :

i) Economic Commission for Asia and the Far East (ECAFE)

ii) The Labour Bureau, Government of India

iii) The State Statistical Bureau of any State

iv) The National Council of Applied Economic Research (NCAER)

Briefly describe the main contents of the publications selected and indicate the usefulness of the data presented. (8+8)=16

3. The Capital-Output Ratio may be defined as the ratio of total capital employed to the value added. What data would you use for estimating capital-output ratios for

i) individual manufacturing industries, and

ii) major sectors of the Indian Economy?

Comment on the reliability of the data available for the purpose. (5+5+6)=16

4. Answer any four of the following questions :

i) Write down the values of national income at factor cost and gross domestic product at market prices for the year 1965-66, given the following data for the same year :

Net domestic product at factor cost	Rs. 21228 crores
Net factor incomes from abroad	Rs. 164 "
Indirect taxes less subsidies	Rs. 2010 "
Allowance for depreciation	Rs. 1199 "

ii) Mention some recent sample surveys of all-India coverage and name the agencies undertaking these surveys.

iii) What are the functions of the National Income Division of the Central Statistical Organisation ?

iv) Is it correct to say that the principal price index numbers in India are mainly urban in character? Give reasons in support of your answer.

v) What source would you use for making an international comparison of the number of primary students per thousand of population? Is the coverage of the information satisfactory?

vi) What are the main activities of the Directorate General of Health Services in the Ministry of Health? (4+4+4+4)=16

Please turn over

/(2)

5. From the official publications furnished for your use, answer any four of the following questions :

[Mention the year of publication, page and table numbers of the publications used in every case without fail. Name the year or years for which figures are supplied and also the base year in case of index numbers.]

i) Obtain the percentages for productions of rice, wheat and coarse cereals (i.e., all other cereals) to the total production of all cereals in India for any recent year.

ii) For any recent year, name the States which account for the largest population of (a) buffaloes and (b) sheep and the largest production of (c) milk and (d) wool.

iii) Write down the number of passengers (originating) and tonnes of freight (carried) by the Indian Railways in any recent year. Furnish also (approximately, if necessary) the average number of kilometers travelled by a passenger in the same year.

iv) Mention three most important sources of revenue and three most important items of expenditure of the Government of India, for a recent year.

v) Present the general index number of wholesale prices and the all-India cost of living index number for any two selected years, so as to indicate the change in the price level in the country during the post-independence period.

vi) Furnish index numbers for any two selected years in recent times showing the growth of :

(a) industrial production of finished steel

and (b) mineral production of bauxite in India. (4+4+4+4)=16

Neatness (2)

GROUP B

(Answer any four questions from this group)

6. (a) Briefly outline the different stages of a statistical investigation.

Explain clearly the concepts of "sampling errors" and "non-sampling errors" in respect of a survey and the methods of minimising them.

(c) the various manual and mechanical tools used now-a-days in processing (4+4+4)=12

7. (a) A binomial distribution is given stating that it will have expected value 10 units and standard deviation 3 units. Examine whether this is possible.

(b) Define a negative binomial distribution. Mention a random experiment and an outcome thereof, for which such a distribution provides a suitable theoretical model.

(c) Show that under certain (limiting) conditions on the parameters (the conditions to be precisely stated by you) a negative binomial distribution can be approximated by a Poisson distribution.
(3+1+3+5)=12

Please turn over

Corrigenda Note

Existing question No.5 of SD Paper I Group A on Official Statistics, is advised to be deleted and replaced by the following revised question, as otherwise it will require handling of Official publications at the examination hall.

Revised Question No.5 :

"Write a brief essay on the role of statistics in national planning."

8.(a) You are given that a certain distribution has cumulants of all orders and all the cumulants of order greater than one are equal to unity. Identify (with reasons) the distribution concerned.

(b) X and Y are standard Normal random variables whose joint distribution is bivariate Normal, the correlation coefficient between them being ρ (Rho). What is the correlation coefficient between the random variables $aX^2 + bY^2$ and $cX^2 + dY^2$ (a, b, c and d, being real numbers) ? (5+7)=12

9.(a) Explain the concept and purposes of an index number. Discuss the significance of the time-reversal and factor-reversal tests. Is there any index number you know of, satisfying both these tests?

(b) What are the different methods known to you for determining the seasonal movements in a time-series? (6+6)=12

10.(a) Write down (without proof) the equations of the regression lines of Y on X and of X on Y, based on a random sample of size n on a bivariate distribution. What is the ratio of the slopes of the two lines?

(b) You are given four points whose rectangular cartesian coordinates are : (1,7), (4,0), (2,4) and (0,3), and also told that the line joining two of them is the regression line of Y on X and the line joining the other two is the regression line of X on Y. Identify the two regression lines if you can, giving reasons for your conclusions.

What are the values of the sample means of the X-values and of the Y-values? (3+1+6+2)=12

11. Suppose you are concerned with a random sample taken on three random variables X_1, X_2 and X_3, having some joint distribution with finite second order moments.

(a) What do you understand by the terms ' Partial correlation coefficient and multiple correlation coefficient in such a context ?

(b) Derive an expression for a partial correlation coefficient in terms of the (total) correlation coefficients.

(c) Show that if r_{ij} $(i,j = 1,2,3,\ i \neq j)$ is the correlation coefficient of X_i and X_j, then

$$r_{23}^2 + r_{31}^2 + r_{12}^2 \leq 1 + 2\, r_{23}\, r_{31}\, r_{12} \qquad (2+2+5+3)=12$$

Neatness (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper II : Probability Theory and Statistical Methods (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks

GROUP A

(Answer any four questions from this group)

1.(a) For any two events A and B show that

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

(b) Write down the extension of the above result to three events A, B and C. (Please note that no proof is required)

(c) If $P(A) = 0.40$, $P(B) = 0.55$, $P(C) = 0.50$;

$P(B \cap C) = 0.25$, $P(C \cap A) = 0.20$, $P(A \cap B) = 0.15$;

and $P(A \cap B \cap C) = 0.05$

find the probability that of the three events A, B, C

i) exactly r occur(s)

ii) at least r occur(s)

iii) at most r occur(s)

for each of the following values $r = 1$, $r = 2$, and $r = 3$. (3+3+3)=9

2.(a) Define probability density function.

(b) For $f(x; \alpha, n) = \begin{cases} c.e^{-\alpha x} x^{n-1}, & x \geq 0 \\ k, & x < 0 \end{cases}$

where α, $n \geq 0$,

determine the constants c and k so that $f(x; \alpha, n)$ is a probability density function.

For what values of α and n does the above correspond to the χ^2 (chi-square) distribution with n degrees of freedom?

(c) Let X_1 and X_2 be independent random variables with probability density functions $f(x; \alpha, n_1)$ and $f(x; \alpha, n_2)$ respectively. Find the distribution of $Y = X_1 + X_2$. Deduce the "reproductive property" of the χ^2 (chi-square) distributions. (1+3+1+3+2)=12

Please turn over

3. For the probability density function

$$f(x_1, x_2) = \frac{1}{8}(x_1^2 - x_2^2)e^{-x_1},$$

$0 < x_1 < \infty$; $-x_1 < x_2 < x_1$;

find :

i) the cumulative distribution function $F(x_1, x_2)$;

ii) the marginal cumulative distribution functions $F_1(x_1)$ and $F_2(x_2)$;

iii) the marginal probability density functions $f_1(x_1)$ and $f_2(x_2)$;

iv) the conditional probability density functions

$$f(x_1 \mid x_2) \text{ and } f(x_2 \mid x_1) \qquad (3+3+3+3)=12$$

4. X_1 and X_2 are independent and identically distributed random variables, and $Y_1 = X_1 + X_2$, $Y_2 = X_1 - X_2$.

i) Assuming that the variance of X_1 is finite, show that Y_1 and Y_2 are uncorrelated

ii) If X_1 and X_2 are uniformly distributed on (0, 1) can Y_1 and Y_2 be considered as independent? Give reasons for your answer. (4+8)=12

5.(a) Write a brief note on the shortcomings of the classical definition of probability.

(b) Carefully state (without proof) the strong law of large numbers (any version), and comment upon the use of the term "strong".

(c) State and prove Chebycheff's inequality. (2+2+3+1+4)=12

Neatness, clarity etc. (2)

GROUP B

(Answer any three questions from this group)

6.(a) Explain with suitable illustrations the criteria of unbiasedness, consistency, efficiency and sufficiency, as used in the theory of estimation.

(b) Let Y_1 and Y_2 be two stochastically independent unbiased estimators for a parameter θ. Assuming that the variance of Y_1 is twice the variance of Y_2, find the constants a and b such that $aY_1 + bY_2$ is an unbiased estimator for θ, with the smallest possible variance for such a linear combination. (10+6)=16

Please turn over

7.(a) Describe the maximum likelihood method of estimation. What are the properties of such an estimator?

(b) Obtain the likelihood-ratio test based on independent random observations $x_1, x_2, \ldots, x_n$ for the hypothesis $\mu = \mu_0$ against the hypothesis $\mu \neq \mu_0$, where $f(x_i) = \frac{1}{\sqrt{8\pi}} \exp\left[-(x_i - \mu)^2/8\right]$

(8+8)=16

8.(a) Give an outline of the Neyman-Pearson theory of testing of hypotheses, explaining clearly the concepts of the errors of type I and type II, power and unbiasedness.

(b) Suppose m and n observations, independent of one another, are available respectively on two normal populations with known means but unknown variances. Write down the statistic which you will use for testing the hypothesis that the ratio of the unknown variances has a specified value (against the alternative that this ratio does not have that specified value).

What can be said about the sampling distribution of this statistic? Obtain the confidence limits to the above ratio with confidence coefficient $(1-\alpha)$. (8+8)=16

9.(a) State the important uses of the 'chi-square' distribution in sampling theory and applications.

(b) State in some detail, the test for independence in a 2 x 2 table. What is Yate's correction for continuity and how is it applied? (8+8)=16

10.(a) Supposing that $r_i (i = 1,2, \ldots, K)$ are the values of the sample coefficients of correlation for K independent random samples of sizes x_i $(i = 1,2, \ldots, K)$ from a bivariate normal population with unknown correlation ρ ρ(Rho), indicate how you would combine the K sample values to get an estimate of ρ (Rho)

(b) State clearly the difference between a parametric and a non-parametric problem. What is meant by the 'robustness' of a statistical procedure? (8+8)=16

11.(a) Explain the terms "linear model" and "linear hypothesis". Write down the (fixed-effects) model for two-way classified data with one observation per cell, stating the assumptions made, and the hypotheses that will be of interest.

(b) Obtain the least-squares estimates of the parameters in the above model. Write down the appropriate analysis of variance table in this case and the statistics that will be used for testing the hypotheses mentioned by you in your answer to 11 (a). (8+8)=16

Neatness (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper III : Sample Surveys and Design & Analysis of Experiments (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks

GROUP A

(Answer any three questions from this group)

1. Comment briefly on the following :

(a) If funds and enumerators are available one should make a complete enumeration - if not, one should take a sample.

(b) For a survey of public opinion in a city on the medium of instruction in schools, a telephone directory is a suitable sampling frame.

(c) For a study of food habits in a town, a cheap and convenient sampling method is to select the homes of students in any one class of any popular school in the town.

(d) If in a simple random sample of shops selected to estimate the total annual sales in an area, a sample shop is found to refuse information at the time of survey, the shop next to it should be surveyed instead. (4+4+4+4)=16

2.(a) Explain how you would locate a random point on a map showing the farms in a village.

(b) If n random points are located on a village map and the farms within which these points lie are selected in the sample (a farm being repeated as many times as the number of random points on it) how will you estimate -

i) the number of cultivated farms in the village

ii) the total area under cultivation

(c) Give the expressions for the estimators of the standard error of estimates (i) and (ii). (2+6+8)=16

3.(a) Explain the method of systematic sampling and discuss its advantages and disadvantages.

(b) Show how systematic sampling is a particular case of cluster sampling and find the condition under which it is more efficient than a simple random sample of the same size, for estimating the population total.

(c) How would you estimate the sampling variance of the estimate of the population total? (5+6+5)=16

4.(a) What is a multi-stage design? Under what circumstances would you use this design?

(b) A simple random sample of n factories is selected and within each sampled factory, a simple random sample of m employees is selected - selection being with replacement at both stages. Show that the variance of the estimate of the mean takes the form

$$V = \frac{1}{n}\sigma_1^2 + \frac{\sigma_2^2}{m}$$

where σ_1^2 and σ_2^2 stand for 'between factory' and 'within factory' variances, respectively.

(c) If the cost function is of the form $C = C_0 + C_1 n + C_2 nm$

find the optimum values of n and m that would minimise V for a fixed cost C. (5+6+5)=16

(M) Please turn over

5. Write brief notes on any three of the following :

i) optimum allocation in stratified sampling

ii) ratio method of estimation

iii) regression estimator

iv) sources of non-sampling errors

v) interpenetrating sub-samples (16)

Neatness (2)

GROUP B

(Answer any three questions from this group)

6.(a) Describe briefly the desirable properties of an experiment and explain how the basic principles of experimentation help you in achieving these.

(b) In a randomised block experiment two treatments turn out to be identical. Obtain expressions for variances of estimated differences between any two treatments. Also write down the analysis of variance table giving expressions for the sums of squares. (8+5+3)=16

7.(a) Explain what is meant by a balanced incomplete block (B.I.B.) design and state without proof, the relations between the parameters of a B.I.B. design. When is a B.I.B. design said to be symmetric?

(b) V treatments are compared in an experiment using a B.I.B. design. Obtain the variance of the estimated difference between the effect of treatment 1 and the average effect of the remaining (V-1) treatments.

(c) Write down the analysis of variance for a simple lattice design, giving expressions for the various sums of squares. Also obtain the expressions for variances of estimates of differences between pairs of treatments. (2+2+1+4+5+2)=16

8.(a) Explain what is meant by the terms :

(i) partial confounding

(ii) balanced partial confounding.

(b) Give a balanced scheme of partial confounding, for studying five factors each at two levels, using five replications in blocks of eight plots each. It is desired that no main effect or two-factor interaction should be confounded in any replication.

[Give complete lay-out for one replication and the key-block for the remaining ones.]

(c) How would you analyse a design using partial confounding in the presence of a concomitant variable? Obtain an expression for variance of the usual estimate for a partially confounded effect.

[If you so choose, you can use your design in (b) above for illustration.] (2+2+6+1+2)=16

Please turn over

9.(a) How would you analyse a Latin Square design, with one row missing?

(b) In a Latin Square design, $T_1 - T_2$ turns out to be the maximum of $T_i - T_j$ for all pairs (i, j) of treatments, whose T_i denotes the total yield for the i-th treatment. Can we use the standard t-test for comparing treatments 1 and 2? If not, explain why it cannot be used and also give an appropriate procedure for comparing the two treatments.

(c) Given the results of an experiment using a Latin Square design, how would you compute the efficiency of this design relative to a randomised block design, using rows as blocks? (6+4+2+4)=16

10. Write short-notes on <u>any two</u> of the following :

i) optimum size and shape of plots (for experimentation)

ii) fractional replications

iii) orthogonal Latin Squares

iv) missing plot technique. (16)

<u>Neatness</u> (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper IV : Applied Statistics (Theoretical)

Time : 4 hours | Full marks : 100

(a) Candidates will be required to answer questions from those two groups of subjects only for which they have registered their option.

(b) Separate answer books are to be used for each of the two groups attempted.

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is not permitted.

GROUP A : ECONOMIC STATISTICS

(Answer any three questions from this group)

1. (a) Explain the factor-reversal test of index number formulae. Which of the standard formulae meets this test?

   (b) Laspeyre's formulae for the price index is said to have an upward bias. Can you say why? (8+8)=16

2. Discuss the difficulties in the time series approach to the estimation of statistical demand functions. How does pooling of cross-section and time series data help in the situation? (12+4)=16

3. Suppose you are fitting a 'Cobb-Douglas production function to cross-section data for firms within a given industry. How would you measure the different variables included in the production function formulation?

   How are the exponents of labour and capital in this production function related to (i) the returns to scale and (ii) the relative shares of labour and capital in production? (5+5+6)=16

4. What is Pareto's law of income distribution? How far is it universal?

   Derive the equations of the Lorenz curve and the Lorenz ratio when the income distribution is Pareto. (3+3+10)=16

5. Explain the concept of national income, distinguishing it from the total of the incomes of all the citizens of a country.

   State briefly how national income is estimated in our country. (8+4+4)=16

6. Write short-notes on any three of the following :

   (a) Choice of trend-type in time series analysis

   (b) The ratio chart and its uses

   (c) Distinction between production function analysis and input-output analysis

   (d) Major objectives of the Five Year Plans of India

   (e) Inferior goods, necessaries and luxuries. (16)

   Neatness (2)

Please turn over

GROUP B : STATISTICAL QUALITY CONTROL

(Answer any three questions from this group)

1.(a) Explain the meaning of statistical control.

(b) A stable process producing 'i' units per unit of time, has μ_0 as its initial setting. Its average level drifting linearly at the rate of θ per unit time, has one sided (upper) specification limit at $\mu_0 + n\sigma$ where n is a constant and σ is the standard deviation of the process.

Assuming that σ remains unchanged and the characteristic follows normal distribution, show that the optimum production run "a" which minimises the expected total loss, is given by the equation :

$$n F(n) + f(n) = \frac{C_r \theta}{\sigma u_i} + n P(T_a) + f(T_a)$$

where C_r = resetting cost

u = loss incurred per defective item

$$T_a = n - \frac{a}{\sigma} \qquad (2+14)=16$$

2. Obtain graphical item by item Sequential Sampling Plan using inspection by attributes. Derive expressions for all the terms used in describing the sample plan, from the following given parameters:

AQL = p_1 Producer's risk = α

LTPD = p_2 Consumer's risk = β

Derive also an expression for the OC function of the plan and hence obtain any five points on the OC curve. (8+8)=16

3. In a life test, it is assumed that the life x of the equipment is distributed as $f(x,\theta) = \frac{1}{\theta} \exp.(-x/\theta)$ x ≥ 0 ($\theta \geq 0$). A random sample of size n is drawn from such a population and put on test. The failed items are not replaced and the test is terminated, when the first r items have failed, where r is a preassigned number.

(a) Derive the distribution of the unbiased estimate of the average life θ.

(b) Obtain an expression for the expected waiting time, to get the first r failures. (12+4)=16

4. A workshop desiring to develop a new product, is interested in conducting an experiment, to determine which method of production would require the minimum processing time.

Four machines and four operators are available for the experimentation. There are also four alternative methods of production.

Suggest a suitable design for the experiment.

Using the usual notations give the estimates of the components of variance. (16)

Neatness (2)

Please turn over

GROUP C : STATISTICAL METHODS IN GENETICS

(Answer any three questions from this group)

1. What are the types of crosses which can provide information on linkage between two factors?

   Assuming that gene-ratios are disturbed for both factors, estimate the values of recombination fraction (p) between the genes and the amount of viability (v), from the backcross data, at both coupling and repulsion phases. Estimate also the standard error of $\hat{p}$. (16)

2. Define the coefficient of heritability under random mating and obtain these coefficients under inbreeding, for 'between inbred lines' and 'within inbred lines'.

   Estimate heritability from intra-sire regression of daughter on dam, stating clearly all the assumptions involved. Mention any other method for estimating heritability. (16)

3. Explain Hardy-Weinberg law of equilibrium under random mating.

   Starting from an initial panmictic population $(p^2, 2pq, q^2)$, in case of the complete positive assortive mating where dominants only mate with dominants and recessives with recessives only, show that the amount of heterozygosity in the nth generation is given by :

   $Y_{n+1} = \frac{Y_0}{1+ng}$, where Y_{n+1} = the heterozygosity in the nth generation and Y_0 = the same in the initial generation. (16)

4. Describe a method of investigating inheritance of M-N blood types from data of parents and children.

   Show how the data on M-N blood-groups in a population, help in examining the extent of inbreeding in a population. (16)

5. Write short-notes on any three of the following :

   i) Concept of 'Repeatability' in mass-selection

   ii) Principle of estimating gene-frequencies from O-A-B blood group system

   iii) Role of selection and mutation in population drift

   iv) Inbreeding and its measurements (16)

   Neatness (2)

GROUP D : VITAL STATISTICS AND DEMOGRAPHY

(Answer any three questions from this group)

1. Give an account of the items on which information should be collected in a population census. Examine the utility of a post-enumeration survey. (16)

2. What are the different measures of reproduction? Write a note on the different factors which influence fertility. (16)

3. (a) Describe King's method of constructing an abridged life-table.

   (b) Discuss the methods of estimating the infant mortality (i.e. the probability of death q_0).

Please turn over

GROUP D : VITAL STATISTICS AND DEMOGRAPHY (Contd.)

4. Write a note on population projections based on (i) the component method and (ii) deterministic models. Discuss their merits.

5. Write short notes on any three of the following :

(a) Stable population;

(b) Working population;

(c) Health surveys;

(d) Migration statistics.

Neatness

GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer any three questions from this group)

1. Write short notes on any four of the following :

(a) Normalized Score

(b) Spearman's Two-Factor Theory

(c) Tetrachoric Correlation Coefficient

(d) Item Validity

(e) Speed and Power Tests

(f) Rotation of Factors (4x4

2.(a) Discuss the effects of group heterogeneity on validity and reliability of a test.

(b) What do you understand by 'correction for guessing'? Briefly discuss the formulae used for correcting the total score in a test. (8

3.(a) What is "item analysis"? Discuss the two indices - (i) difficulty and (ii) discrimination, that could be obtained for each item.

(b) Write a short note on the effect of test length on the reliability and validity. (8

4.(a) What do you understand by 'standard score'? Why is it necessary to convert raw scores to standard scores? Outline the steps that you would follow to calculate percentile scores.

(b) Describe the standard error of measurement of a test? Under what conditions, will the standard error of measurement of a test be equal to zero? (8

5. Describe the Centroid method of estimating the factor loadings of the tests included in a battery, stating the underlying assumptions What are factor scores? How would you estimate them?

Neatness

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper V : Methods of Numerical Computation; Descriptive Statistics; and Official Statistics (Practical)

Time : 5 hours Full marks : 100

(i) Figures in the margin indicate full marks

(ii) Use of calculating machines is permitted.

GROUP A

(Answer any two questions from this group)

1.(a) The following data relate to the domestic demand for motor fuel, in U.S.A. Estimate the missing figures for 1925 and 1927.

Year	Domestic demand for motor fuel (in thousand barrels)
(1)	(2)
1922	10,639
1923	13,062
1924	15,417
1926	21,818
1928	27,374

(b) The population of a district in Bihar in five consecutive groups of age last birthday, is given below. Obtain the population at age 42, last birthday.

Age (last birthday)	Population (units)
(1)	(2)
30 - 34	148, 632
35 - 39	132, 571
40 - 44	119, 870
45 - 49	101, 532
50 - 54	98, 568

(6+7)=13

2.(a) AB is the base of a semi-circle centre O and radius unity. The points P and Q bisect OA and OB respectively. The area between the base PQ, the semi-circle and the ordinates at P and Q is $\frac{\pi}{6} + \frac{\sqrt{3}}{4}$. Use this result to find the approximate value of π to 3 places of decimals by using a quadrature formula.

(b) Calculate the real root of the equation $x^3 + x - 3 = 0$ by inverse interpolation, correct to 3 places of decimals. (7+6)=13

Please turn over

3.(a) Given,

Age x	l_x
50	73, 499
51	72, 729
52	71, 753
53	70, 809
54	69, 154

estimate $\frac{dl_x}{dx}$ and $\frac{d^2 l_x}{dx^2}$ at x = 50

(b) Find the inverse of the following matrix :

$$A = \begin{vmatrix} 1.234 & 0.005678 & -235.6 \\ 67.82 & -2341.0 & 1.234 \\ 0.9876 & 87.65 & -56.78 \end{vmatrix}$$ (7+6)=13

GROUP B

(Answer all the questions from this group)

4. The following table shows the birth and death rates per 1000 people in U.S.A. for the years 1915-55. Give graphical representation for the same

Year	1915	1920	1925	1930	1935	1940	1945	1950	1955
Birth rate	25.0	23.7	21.3	18.9	17.9	19.5	23.6	24.6	25.0
Death rate	13.2	13.0	11.7	11.3	10.8	10.6	9.6	9.3	9.0

(8)

5. A skilled typist, on routine work, kept a record of mistakes made per day during 300 working days.

Mistake per day	0	1	2	3	4	5	6
Number of days	143	90	42	12	0	3	1

Compute the frequencies of the Poisson series which has the same total frequency and mean as the above distribution. (10)

6. EITHER

Certain mass-produced articles of which 0.5 percent are defective, are packed in cartons each containing 100 articles.

What proportion of cartons is free from defective articles, and what proportion contains 2 or more defectives? (12)

OR

The following table gives the frequencies of scores between certain limits :

Scores	Frequency
Less than 40	30
40 or more but less than 50	33
50 or more	37

Assuming the distributions to be normal calculate the number of candidates passing in the 1st, 2nd and 3rd divisions.

Minimum scores for the different divisions are as follows :

Division of pass	1	2	3
Minimum score	60	45	36

7. EITHER

In an experiment on wheat, fertilizers were applied at various levels with resulting yields as follows :

Fertilizer Level(x)	0	5	10	15	20	25	30	40	60
Yield(y)	26.2	31.1	34.0	35.3	36.3	37.1	37.8	38.6	38.0

Fit the equation $y = 39.0 - \exp(a + bx)$, to the given data.

Plot the observed data and draw the fitted curve on the same graph paper. (20)

OR

The numbers of letters posted in a certain city on each day in a period of five consecutive weeks are given below :

Serial No. of week t	Number of letters (in hundreds) posted on different days of the week						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
(1)	(2)	(3)	(3)	(4)	(5)	(6)	(7)
1	18	161	170	154	143	161	76
2	18	165	179	157	168	195	85
3	21	162	169	153	139	185	82
4	24	171	181	170	162	179	93
5	27	172	196	180	173	202	120

i) Fit a straight line trend $y = a + bt$ to the weekly totals (y) and use the same to obtain the trend value for each day.

ii) Find the ratios of the daily figures to the corresponding trend values. Using the median as the representative figure, calculate indices of variation from day to day within a week.

GROUP C

(Attempt both the questions from this group)

8.(a) Discuss the evolution of official statistics on the production of manufacturing industries in India, stating the names of official publications, authorities publishing them, method of collection of data and their reliability.

(b) Comment on the growth of industrial production in India by collecting suitable data for the latest ten available years. (6+6)=12

9. Collect, from the official publications furnished, the following data for the latest five available years.

(a) The acreage and production of rice for the States, Andhra, Bihar, Bombay, Orissa and West Bengal.

(b) Number of passengers killed in railway accidents and number of railway servants killed in railway accidents, in Indian Railways. (6+6)=12

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper VI : Statistical Methods; Design & Analysis of Experiments and Sample Surveys (Practical)

Time : 5 hours Full marks : 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

GROUP A

(Answer any two questions from this group)

1.(a) For 7 newly born babies the weights at birth as well as at the end of a month, are given below :

Serial no.	Weight at birth (Kilograms)	Weight after a month (Kilograms)	Serial no.	Weight at birth (Kilograms)	Weight after a month (Kilograms)
(1)	(2)	(3)	(1)	(2)	(3)
1	4.23	5.58	5	3.87	5.00
2	3.04	5.29	6	4.39	5.60
3	4.90	5.97	7	4.80	5.75
4	3.62	5.06			

Test whether the weights of babies at birth and after a month are equally variable.

(b) From the results of a crop cutting experiment carried out in a certain district divided into 6 blocks, the following correlation coefficients between weights of ears of wheat as harvested and weights of grains of wheat as available after threshing and cleaning, were obtained :

	Blocks					
	1	2	3	4	5	6
Sample size	274	179	233	84	255	50
Correlation coefficient	.9055	.8710	.9210	.9365	.9508	.9780

i) Test whether the correlations differ from block to block.

ii) Assuming the correlations do not differ, obtain also a combined estimate of the correlation coefficient for the entire district. (20)

2.(a) A student obtained the following figures as a random sample of size 15, from a normal population with mean 4 and S.D. 2. Use Kolmogorov-Smirnov goodness of fit statistic, to test the randomness of the sample.

2.32, 6.74, 3.61, 4.70, 9.64
8.24, 2.02, 4.68, 5.28, 4.22
3.06, 2.28, 4.02, 3.91, 3.18

(b) The following table shows 42 children classified according to the nature of their teeth and type of feeding :

	Normal teeth	Malformed teeth	Totals
Breast fed	4	16	20
Bottle fed	1	21	22
Total	5	37	42

Apply Fisher's exact probability test, to judge whether there is any association between breast feeding and normal teeth. (20)

Please turn over

3.(a) An experiment was performed to compare the effects of three treatments on the yield of a plant which is grown in pots. 5 pots were chosen randomly from a large lot of pots and in each 6 seeds were planted, 2 of which were subjected to treatment A, the other 2 to treatment B, and the remaining 2 to treatment C. The yields from the plants grown are given below. Analyse the data for the purpose in view.

Pot Number	Treatment					
	A		B		C	
	Plant I	Plant II	Plant I	Plant II	Plant I	Plant II
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	15	16	20	22	25	24
2	20	25	30	29	40	35
3	16	14	18	16	20	18
4	18	15	20	22	22	26
5	21	22	20	19	30	28

(b) A new sealing device for electric bulbs is to be recommended if 10% or less defectives result from it and it is to be declared as unworkable if 30% or more defectives result from it. An experiment is to be carried out Sequentially to judge the device. It is desired that the risk of declaring the device as unworkable when it is actually recommendable, should not be more than .02 and the risk of recommending an unworkable device should not be more than .03. Draw a chart for applying the Sequential probability test under this situation.

Data below represent the results of observations on 22 bulbs sealed by the new device. Plot the results on your chart, and see whether

i) the experiment has been over continued, or
ii) it still requires to be continued, or
iii) it is time to stop right now.

In case the experiment need not be continued further, give your decision.

Serial number of bulb	Result of observation (T = True and D = Defective)	Serial number of bulb	Result of observation (T = True and D = Defective)
(1)	(2)	(1)	(2)
1	T	13	T
2	T	14	D
3	D	15	T
4	T	16	T
5	T	17	D
6	T	18	T
7	T	19	T
8	T	20	T
9	D	21	T
10	T	22	D
11	D		
12	D		

Please turn over

GROUP B

(Answer any two questions from this group)

4. The data given below relate to yield in ounces for each plot of a randomised block experiment, involving 7 strains of 'mung' - a variety of pulse.

Strain →		1	2	3	4	5	6	7
Block	I	18	27	22	20	21	25	15
	II	20	15	21	11	22	27	15
	III	15	21	14	19	00	22	16
	IV	32	20	25	19	25	29	18
	V	20	25	26	24	20	32	14

The experimenter reports that the plot allotted to strain 5 in Block III had been badly damaged by birds etc., thus depressing the yield of this plot considerably.

Analyse the data treating this plot as missing and obtain the standard errors of the various type of strain differences. (10+5)=15

5. The following table gives the results of a double Lattice experiment. Carry out the intra-block analysis of the data and obtain,

i) the standard errors of the two different types of treatment comparisons, and

ii) the average variance of all the treatment comparisons.

Replication I - (figures in brackets indicate treatment numbers)

Block 1	(1) 14.2	(2) 16.0	(3) 14.8	(4) 16.7
Block 2	(5) 14.6	(6) 15.8	(7) 16.0	(8) 14.9
Block 3	(9) 17.3	(10) 19.8	(11) 19.4	(12) 10.3
Block 4	(13) 16.0	(14) 13.6	(15) 15.3	(16) 11.6

Replication II - (figures in brackets indicate treatment numbers)

Block 5	(1) 17.3	(5) 15.0	(9) 19.5	(13) 19.6
Block 6	(2) 18.7	(6) 18.8	(10) 20.2	(14) 19.3
Block 7	(3) 16.4	(7) 23.2	(11) 19.1	(15) 16.5
Block 8	(4) 18.0	(8) 22.8	(12) 15.5	(16) 13.2

You are furnished with the total sum of uncorrected squares as being equal to 9354.72. (9+3+3)=15

Please turn over

6. The following table gives the lay-out plan and yields of a 2^4 field experiment on Beans, conducted by Rothamsted Experimental Station in 1936. The factors were :

Dung (D): none; 10 tons per acre
Nitrochalk (N): none; 0.4 cwt. N per acre
Superphosphate (P) : none; 0.6 cwt. P_2O_5 per acre
Muriate of potash (K): none; 1.0 cwt. K_2O per acre

Replication 1				Replication 2			
Block number				Block number			
1a		1b		2a		2b	
treat-ment	yield	treat ment	yield	treat-ment	yield	treat-ment	yield
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
p	43	dp	50	npk	43	nk	43
k	55	nk	44	d	42	dp	52
d	53	dk	43	p	39	(1)	57
npk	36	pk	51	dnk	34	np	39
dnk	41	dnpk	44	n	47	pk	56
dnp	48	(1)	58	dnp	52	dk	52
dpk	55	dn	41	k	50	dpnk	54
n	42	np	50	dpk	44	dn	42

i) Identify the confounded effect or effects and analyse the data.

ii) Test if the main effect of nitrochalk (N), and the interaction of Dung and Superphosphate (DP) are significant. (2+9+2+2)=15

GROUP C

(Answer any one question from this group)

7. Using the data given below, compare the precisions attained in the case of simple random sampling with replacement,

i) unstratified
ii) stratified with proportional allocation
iii) stratified with optimum allocation,

when the average production of the Saw mills is estimated from a sample of 1000 mills:

stratum number	annual production (M bd ft.)	number of Saw mills	average production	standard deviation
(1)	(2)	(3)	(4)	(5)
1	5000 and over	540	11030	9000
2	1000 - 4999	4800	1780	1200
3	Under 1000	30950	204	300

If the cost of collecting each schedule is

(a) Re.1/- for mills 5000 and over M bd ft,

(b) Re.4/- for mills producing 1000-4999 M bd ft. and

(c) Re.9/- for mills under 1000 M bd ft.,

what is the total expenditure for the optimum allocation in (iii) above?

Using the same total expenditure, what allocation would give you the smallest error, taking account of costs?

What will be the total sample size, with this allocation? (3

Please turn over

8. The following table shows the number of inhabitants (in thousands) in each of a simple random sample of 49 cities, drawn without replacement, from the population of 196 large cities, as obtained in 1920 and 1930.

The true 1920 total X is 22,819(000). Compute the regression estimate of 1930 Y total number of inhabitants, in the 196 large cities.

Find the approximate standard error of this estimate and compare its precision with that of the ratio estimate.

Sizes of 49 sample cities in the years 1920 and 1930

(Figures in thousands)

1920 (x_i)	1930 (y_i)	1920 (x_i)	1930 (y_i)	1920 (x_i)	1930 (y_i)	1920 (x_i)	1930 (y_i)
(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
76	80	78	106	64	77	256	288
138	143	66	86	56	142	43	61
67	67	60	57	40	60	25	57
29	50	46	65	40	64	91	85
381	464	2	50	38	52	43	50
23	48	507	634	136	139	208	317
37	63	179	260	116	130	36	46
120	115	121	113	46	53	161	232
61	69	50	64	243	291	74	93
387	459	44	58	87	105	45	53
93	104	77	89	30	111	36	54
172	183	64	63	71	79	50	58
						48	75

(30)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper VII: Applied Statistics (Practical)

3:5 hours Full marks:100

(a) Candidates will be required to answer questions from those two groups of subjects only, for which they have registered their options.

(b) Separate answer books are to be used for each of these two groups attempted.

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is permitted.

GROUP A: ECONOMIC STATISTICS

(Answer all questions from this group)

EITHER

The following table shows the average per capita consumption of cereals and prices of cereals in rural India, observed during four different periods P_1, P_2, P_3 and P_4

(Period 1 - July 1954 to March 1955; Period 2 - May 1955 to November 1955; Period 3 - September 1957 to May 1958; and Period 4 - July 1959 to June 1960)

	Consumption per 30 days (seer)				prices (paisa per seer)			
	P_1	P_2	P_3	P_4	P_1	P_2	P_3	P_4
ce|	8.69	7.94	7.58	8.22	37	39	52	52
eat	1.51	2.73	2.43	2.78	37	32	41	44
arse ereals	7.66	8.19	7.75	7.71.	22	20	31	36

Calculate a price index number of cereals for July 1959-June 1960, taking July 1954-March 1955 as base; adopting the chain base system and using the Laspeyre's formula at each step. [15]

OR

Plot the following series of American wholesale price index of all commodities (1926 = 100) and determine the trend by using an appropriate moving average formula:

Year	Price Index	Year	Price Index	Year	Price Index
1890	56.2	1900	56.1	1910	70.4
1891	55.8	1901	55.3	1911	64.9
1892	52.2	1902	58.9	1912	69.1
1893	53.4	1903	59.6	1913	69.8
1894	47.9	1904	59.7	1914	68.1
1895	48.8	1905	60.1	1915	69.5
1896	46.5	1906	61.8	1916	85.5
1897	46.6	1907	65.2	1917	117.5
1898	48.5	1908	62.9	1918	131.3
1899	52.5	1909	67.6	1919	138.6

[15]

Estimate the Engel elasticity of consumption of edible oil from the following data based on an enquiry in rural India.

Please Turn Over

1. class ranges of per person monthly consumer expenditure on all items (Rs.)	0- 8	8-11	11-15	15-21	21-28	28-43	43-55	55.
2. percent of population	15.5	17.8	22.3	19.6	12.7	8.5	2.0	1.
3. per person monthly consumer expenditure (Rs.) on all items (x):	6.2	9.6	12.8	17.5	24.0	33.1	49.0	89.
4. per person monthly consumer expenditure (Rs.) on edible oil (y):	0.12	0.22	0.33	0.42	0.57	0.83	- 1.34	3.

You may assume the constant elasticity form $y = ax^{\circ}$ of the Engel curve.

3. EITHER

The following table gives the age-distribution of fathers of boys born alive in Norway, during 1871-1900:

age (years)	no. of boys	age (years)	no. of boys
- 20	2217	45 - 50	38916
20 - 25	36147	50 - 55	17218
25 - 30	94272	55 - 60	6492
30 - 35	112670	60 - 65	2571
35 - 40	95965	65 - 70	952
40 - 45	69714	70 -	399

The mean of the distribution is 35.699 and the standard deviation 8.410.

Estimate the parameters of the lognormal distribution which may be fitted to the given data, by any suitable method, and calculate the expected frequencies for the classes 0-20, 35-40 and 70 or more.

OR

Given below is the transactions matrix of a certain economy, in arbitrary value units. Calculate the supply requirements for the four sectors if the final demand vectors were (50,120, 60, 280) instead of (60, 105, 40, 320).

Producing Sectors	Using Sectors S	A	B	F	Final Demand	Total Supply
Services (S)	20	25	15	80	60	200
Agriculture (A)		25		120	105	250
Basic Industry (B)		25	45	40	40	150
Finished Goods (F)				80	320	400

GROUP B: STATISTICAL QUALITY CONTROL

(Answer any two questions from this group)

11. From an electronic factory producing a resistive component, 25 samples of size 4 were taken and the resistance of each component was recorded in Ohms. The specification was given as '285 Ohms maximum' (upper limit only). The samples were taken every quarter of an hour and it is known that if the process is given a fresh setting, it will produce components, initially, with a mean $\mu = 250$ Ohms.

Please Turn Over

(contd.)

Analyse the data and recommend a suitable control chart device to the factory management.

Sample No.	Resistance of sampled components (Ohms) S_1	S_2	S_3	S_4	Average
1	252	262	261	247	255.5
2	250	256	257	255	254.5
3	257	245	252	265	254.8
4	254	252	255	250	253.0
5	252	257	252	232	248.3
6	227	264	238	253	245.5
7	260	261	241	261	255.8
8	261	252	255	257	256.3
9	256	262	260	255	258.3
10	257	272	245	255	257.3
11	260	265	265	263	263.3
12	260	258	257	260	258.8
13	245	262	240	270	254.3
14	205	270	270	244	247.3
15	246	270	277	271	266.0
16	260	267	275	275	269.3
17	270	271	281	263	271.3
18	252	265	270	268	263.8
19	260	277	273	270	270.0
20	272	275	270	283	275.0
21	275	275	257	257	270.5
22	267	265	283	277	273.0
23	277	273	275	273	274.0
24	277	277	302	271	281.8
25	274	240	280	289	270.8

[25]

a) A company which is developing a new product makes 25 items on a trial basis, 4 of these were found to be defective. Some alterations in the design were then made and n items produced of which none were found to be defective. How large should be n so that the new design could be considered an improvement at 1 per cent level of significance?

b) From the Dodge-Romig Sampling Inspection Tables, select plans to satisfy:

i) lot size: 2500, process average: 0.5 per cent; AOQL: 2.50 per cent; type of sampling: double

ii) lot size: 2600; LTPD: 4.0 per cent; process average: 0.7 per cent; type of sampling: single.

c) For each of the plans in (b) above, obtain the approximate probabilities of acceptance (OC) at the following values of incoming quality 0.6 per cent, 3 per cent and 8 per cent.

[9+6+10]=[25]

The following table gives the results of an experiment conducted by a factory to study the effect of three factors (method of processing, temperature and raw material) on the yield of a process. The process was carried out by each of the two methods M_1 and M_2; at each of the three temperatures T_1, T_2 and T_3; and four batches of raw material B_1, B_2, B_3 and B_4 were used. Two trials were carried out with each set. The experimental data are given below.

Please Turn Over

3.(contd.) Analyse the data given below. What specific suggestions can you make to increase the yield?

Temperatures	Batch$_1$		Batch$_2$		Batch$_3$		Batch$_4$	
	M_1	M_2	M_1	M_2	M_1	M_2	M_1	M_2
T_1	76.1	80.5	68.6	73.2	67.1	80.2	74.4	77.5
	74.0	79.3	69.4	76.0	71.4	72.9	78.6	75.3
T_2	74.0	80.9	70.0	74.4	77.6	76.8	70.2	75.6
	76.1	80.2	73.0	79.3	75.4	80.2	74.5	80.3
T_3	74.0	78.2	74.8	81.0	73.6	80.9	78.2	85.8
	82.9	83.8	74.9	79.0	72.8	77.8	77.6	79.6

[2

GROUP C: STATISTICAL METHODS IN GENETICS

(Answer any two questions from this group)

1. In a replicated progeny row trial on a fibre crop, with a compact family block layout, there are three families, each consisting of three progenies. The families were allocated to main plots and the progenies to sub-plots. Data on fibre percentages after baling are as shown below for each plot:

Data on Fibre percentages:

Family	Progeny	I	II	III	IV	V
A	1	31	29	28	29	30
	2	30	29	29	31	28
	3	29	28	30	27	29
B	1	35	31	33	33	30
	2	34	36	35	36	35
	3	32	30	36	35	33
C	1	28	27	31	29	27
	2	30	28	27	29	29
	3	28	29	30	28	30

1.a)

After using Bartlett's test of homogeneity of three 'within family error-variances', complete the pooled ANOVA table. Test the differences in two progeny means 'within the same family' and also 'belonging to different families', after calculating the appropriate standard errors of differences. Rank the progenies according to their fibre percentages behaviour.

[2 b)

2. In a two factor segregation, factor A-a showing complete dominance of A over a and factor B-b showing no dominance, are known to be linked. The progeny from a cross $(\frac{AB}{ab} \times \frac{AB}{ab})$ showed the following segregation:

Phenotypic classes:

	$\frac{A(BB)}{24}$	$\frac{A(Bb)}{30}$	$\frac{A(bb)}{5}$	$\frac{a(BB)}{8}$	$\frac{a(Bb)}{36}$	$\frac{a(bb)}{28}$
Frequency						

Test whether the Mendelian law of segregation is followed. Estimate the linkage by a suitable method.

Assuming no sex-differential in respect of recombination, obtain an efficient estimate of the recombination fraction and its standard error.

[

3. The amount of Dutch clover in a forage stand was estimated by a mechanical counter (x_1) and by eye (x_2). The two treatments to be discriminated were randomized in 15 blocks of two plots each, so that 14 degrees of freedom could be taken out for block differences, giving an analysis of variance set up as appears in next page :

Please turn over

(mm)

S.S. = Sum of squares
S.P. = Sum of products

Sources	d.f.	S.S.(x_1)	S.S.(x_2)	S.P.($x_1 x_2$)
Between population	1	13.47	8.43	10.65
Between blocks	14	93.11	54.69	60.95
Within population	14	20.44	6.41	4.89
Total	29	127.02	69.53	76.49

Obtain the best discriminant function (y) and the corresponding ANOVA.

If instead, a discriminant based on the sum of the variables (i.e. $z = x_1 + x_2$) is set up, find out the relative efficiency of selection. Also work out the correlation coefficient between these two discriminants (viz. y and z). [23]

Neatness [2]

GROUP D: VITAL STATISTICS AND DEMOGRAPHY

(Answer any two questions from this group)

a) From the data given in the following table find percentage reduction below potential due to (i) contraception, and (ii) contraception and subfecundability combined. Interpret the results.

Estimated reduction of actual below potential births per couple, due to 'contraceptions' and to 'subfecundability and contraception combined' by duration of marriage (i.e. years married).

Type of data	Years married			
	under 5	5-9	10-14	15 or more
Number of couples	649	869	686	509
Average number of births:				
Potential	1.43	3.99	6.32	9.16
Observed, fecund couples	1.00	2.34	2.99	3.52
Observed, all couples	0.91	2.03	2.54	2.91

b) Calculate the general fertility rate and the gross reproduction rate from the following data, assuming that for every 100 girls 106 boys are born.

Age of women	Number of women	Age-specific fertility rate
15 - 19	212,619	98.0
20 - 24	198,732	169.6
25 - 29	162,800	158.2
30 - 34	145,362	139.7
35 - 39	128,109	98.6
40 - 44	106,211	42.8
45 - 49	86,753	16.9

(25)

2.a) From the following values of the usual life-table functions, calculate the expectation of life at 30.

Age group (X to X+n)	$_nq_x$	
25 - 30	.0284	l_{25} = 87,870
30 - 35	.0352	T_{45} = 1,686,581
35 - 40	.0501	
40 - 45	.0606	
45 - 50	.0872	

State the approximations you have to use.

b) Confining attention to persons below 25 only find out the standardised mortality ratio for the two places A and B from the following data:

Age	Standard population	Mortality rates per 1000	
		A	B
Under 1 year	20,883	26.884	43.182
1 - 4	86,376	1.027	2.136
5 - 14	161,375	0.400	0.786
15 - 24	146,641	0.801	1.907

(25)

3. The following table gives the distribution of married couples classified by age at return marriage, for the period from 1961 to 1966 in a recent survey. Find the lines of regressions and correlation coefficient between the ages of husband and wife. Interpret the results.

Age of husband at return marriage	Age of wife at return marriage								Total
	13-14	15-16	17-18	19-20	21-22	23-24	25-26	27-28	
15 - 16	3	-	-	-	-	-	-	-	3
17 - 18	2	5	1	-	-	-	-	-	8
19 - 20	4	7	10	2	-	-	-	-	23
21 - 22	1	7	20	10	1	-	-	-	39
23 - 24	-	6	12	14	9	3	-	-	44
25 - 26	-	3	13	8	5	3	2	-	34
27 - 28	-	1	7	5	6	1	-	-	20
29 - 30	-	-	4	2	5	1	1	-	13
31 - 32	-	-	1	1	7	1	2	1	13
33 - 34	-	-	-	2	1	1	-	-	4
35 - 36	-	-	-	-	1	1	2	1	5
Total	10	29	68	44	35	11	7	2	206

(25)

GROUP E: EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer all questions from this group)

A test contains 25 items and its mean, standard deviation and reliability are 17.3, 3.8, and .86 respectively. Estimate the variance and reliability if 75 items are added to it. [12]

The following information on a test is available. Find the test reliability by using (a) difficulty values of the items and standard deviations of the total score, (b) mean and standard deviations of the total scores.

Item No.	Difficulty Value	Item No.	Difficulty Value	Item No.	Difficulty Value
1	.70	8	.86	15	.32
2	.60	9	.74	16	.30
3	.65	10	.36	17	.24
4	.75	11	.46	18	.26
5	.48	12	.44	19	.25
6	.42	13	.52	20	.20
7	.35	14	.48		

The standard deviation of the total score on the test is 4.5. [13]

Using the intercorrelation matrix given below, compute the factor loadings of the first two factors using the Centroid method. (Guess the communalities by using the maxim maximum value in the corresponding column and proceed.)

Variables	1	2	3	4	5	6
1	-	.63	.00	.00	.27	.00
2	.63	-	.32	.36	.21	.00
3	.00	.32	-	.72	.27	.24
4	.00	.36	.72	-	.00	.00
5	.27	.21	.27	.00	-	.72
6	.00	.00	.24	.00	.72	-

[25]

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper VIII : Subjects of Specialisation - (I)

Time 4 hours Full marks 100

(a) Candidates will be required to answer questions from that group for which they have registered their options.

(b) Figures in the margin indicate full marks.

Group A: Economic Statistics (Econometrics).

(Answer any five questions from this group)

1. (a) If the utility function is

$U = \frac{1}{2}(a_{11} x_1^2 + 2 a_{12} x_1 x_2 + a_{22} x_2^2)$, show that the demands are given in terms of prices (p_1, p_2) and income (M) by

$$a_{11} x_1 + a_{12} x_2 = p_1$$

$$a_{12} x_1 + a_{22} x_2 = p_2$$

and $p_1 x_1 + p_2 x_2 = M$

(b) Deduce that the demands are linear in M but not in p_1 and p_2.

(c) Show that for stability, $a_{22} p_1^2 - 2 a_{12} p_1 p_2 + a_{11} p_2^2 < 0$ for any p_1 and p_2. Deduce that $a_{11} < 0$, $a_{22} < 0$ and $a_{12}^2 < a_{11} a_{22}$.

Check that the demand curves are then downward sloping.

(8+4+8)=20.

2. (a) For a single commodity, let increasing α represent an upward shift in demand: $D(P, \alpha)$ with $\frac{\partial D}{\partial \alpha} > 0$. Show that the variations in equilibrium price and purchases are:

$$\frac{dP}{d\alpha} = \frac{\partial D}{\partial \alpha} \times \frac{1}{\frac{\partial}{\partial P}(S - D)}$$

and $$\frac{dX}{d\alpha} = \frac{\partial D}{\partial \alpha} \times \frac{\frac{dS}{dP}}{\frac{\partial}{\partial P}(S - D)}$$

Deduce that the upward shift in demand, if equilbrium is stable, raises prices, and it raises purchases only if S is upward sloping.

2. (b) There are three markets for commodities (apart from m raire) and there is an upward shift in the supply of the fir

$\frac{\partial}{\partial \alpha}(Y_1 - X_1) > 0$, show that

$$a_{11}\frac{dp_1}{d\alpha} + a_{12}\frac{dp_2}{d\alpha} + a_{13}\frac{dp_3}{d\alpha} = -\frac{d}{d\alpha}(Y_1 - X_1)$$

$$a_{21}\frac{dp_1}{d\alpha} + a_{22}\frac{dp_2}{d\alpha} + a_{23}\frac{dp_3}{d\alpha} = 0$$

$$a_{31}\frac{dp_1}{d\alpha} + a_{32}\frac{dp_2}{d\alpha} + a_{33}\frac{dp_3}{d\alpha} = 0$$

where $Y_r - X_r$ represent excess of supply over demand of commodity r, and $a_{rs} = \frac{\partial}{\partial p_s}(Y_r - X_r)$.

Solve for the three price variations.

(10+10

3. Let induced investment (I) be given by $I_t = v(Y_{t-1} -$ where Y_t is the income at period t and v, a constant. Furthe let consumption (C) be given by $C_t = c_1 Y_{t-1} + c_2 Y_{t+2}$ (c_1, c_2 constants). And further, let investment and savings equal ex post.

Determine the equilibrium path of Y_t. What will t shape of the path that Y_t will follow after an initial distu bance?

Critically discuss the various alternatives. Will the alternatives depend on the magnitude of the disturbance?

4. (a) Explain the problem of multicollinearity, and develo technique of dealing with it. Is it always possible to get this difficulty by suitable estimation procedures?

(b) Suppose that two explanatory variables are exactly r according to $Z_{2t} = 3.0 + 0.5\, Z_{1t}$, and that

$$Y_t = -1.8 + 2.0\, Z_{1t} - 0.7\, Z_{2t} + v_t.$$

Find an equation for Y_t that is equivalent to the f going and that has an arbitrary value α for the coefficie of Z_{2t}.

What are the numerical coefficients if α is set (i) (ii) at 1.5?

(10+10

5. Describe mathematically the mechanism of market equilibrium under the conditions of perfect competition. (20)

6. Consider the value of transaction matrix $V = [V_{rs}]$ in a closed transaction economy, where fixed technical coefficients $a_{rs} = \frac{x_{rs}}{X_s}$ are defined in quantity terms.

Show that $V = \hat{P} [I - A] \hat{X}$ where matrix $A = [a_{rs}]$ and $\hat{P}$ and $\hat{X}$ are the diagonal matrices from the price vector P and the quantity vector X respectively. Deduce the conditions $[I - A] X = 0$ and $P' [I - A] = 0$.

Let A^* be the matrix of order $(n-1)x(n-1)$ derived from A by omitting its n-th row and nth column, and let X^* and P^* be $(n-1)(-)$ vectors derived from vectors X and P after omitting their n-th terms. How will you interpret the vectors $(I - A^*)X^*$ and $P^* (I - A^*)$, if the n-th activity, which is omitted, is interpreted as households? (20)

7. Describe an Input-Output table, explaining how the incomes accruing to various sectors of the economy (household, government and enterprises) are shown in it. Can expenditure incurred by different sectors be also obtained from the table? How transactions with the 'Rest of the World' are depicted in it? (20)

8. Examine the identifiability of the β and γ parameters in the following models:

(a) $y_{1t} + \beta_{12} y_{2t} + \gamma_{11} x_{1t} = u_{1t}$

$\beta_{21} y_{1t} + y_{2t} + \gamma_{22} x_{2t} = u_{2t}$

where variance-covariance matrix of disturbance is

$$\varphi = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

(b) $y_{1t} + \beta_{12} y_{2t} = u_{1t}$

$\beta_{21} y_{1t} + y_{2t} + \gamma_{21} x_{1t} = u_{2t}$

where variance-covariance matrix of disturbance, is

$$\varphi = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix}$$

(c) $y_{1t} + \beta_{12} y_{2t} = u_{1t}$

$\beta_{21} y_{1t} + y_{2t} + \quad_{21} x_{1t} = u_{2t}$, but where variance-covariance matrix of disturbance, is

$$\varphi = \begin{bmatrix} \sigma_{11} & 0 \\ 0 & \sigma_{22} \end{bmatrix} \qquad (2$$

Group B: Techno-commercial Statistics - (Statistical Quality Control).

(Answer any four questions from this group)

1. (a) Distinguish between chance and assignable variations in the quality of a manufactured product.

(b) When is a manufacturing process said to be in a state of statistical control?

(c) How do you construct a control chart for fraction defective, when the numbers of units inspected in different samples are not the same?

(6+5+14) = 2

2. In what respects, a cumulative sum control chart for sample means, differs from (i) an ordinary $\bar{x}$-chart and (ii) a chart for moving averages?

How is a cumulative sum control chart for sample means related to sequential tests for the mean of a normal population?

Outline the steps in using such a chart.

(8+12+5) = 2

3. (a) Compare ($\bar{x}$, R) charts with analysis of variance as tool for detecting shifts in the process mean.

(b) Examine the uses of range and mid-range in quality control work.

(15+10)= 2

4. (a) What considerations should influence the selection of an acceptance sampling plan?

(b) Explain the term 'A.O.Q.L.'. Derive its expression for single sampling inspection plan, for acceptance-rectification purposes.

(10+15)= 2

5. (a) When do you prefer A.O.Q.L. sampling plans to L.T.P.D. sampling plans?

(b) Indicate advantages of sequential sampling plans over single sampling plans for acceptance inspection.

(c) Describe briefly a sampling inspection plan for continuous production; state its demerits, if any.

(5+8+12) = 25

6. Write notes on any four of the following:

(i) Uses of work measurement data

(ii) Advantages of factorial experiments over one-factor experiments

(iii) Measurement of Process Capability

(iv) Control chart for the weighted total of defects in a complex assembly

(v) Specification limits, tolerance limits and control limits

(vi) Control chart for extreme values. (25)

Group C: BIOMETRIC METHODS (I)
No candidate available

Group D: Design & Analysis of Experiments (I)

(Answer any five questions from this group)

1. (a) State clearly the assumptions underlying the technique of analysis of covariance.

(b) How would you decide whether observations on concomitant variables have increased the efficiency of an experiment?

(c) Sketch the statistical analysis of a Latin square design when one observation on each of two concomitant variables is available on every plot.

(5+5+10) = 20

2. Let v treatments be arranged in b blocks such that (i) any treatment occurs at most once in any block, (ii) the i-th treatment is replicated r_i times (i = 1, ..., v), (iii) the size of the j-th block is k_j, (j = 1, ..., b).

Let $N = (n_{ij})$ be the incidence matrix of the design.

(a) Obtain the analysis of variance of this design under the usual assumptions

(b) Define the C-matrix

(c) When is this design called (i) connected, (ii) balanced?

(d) Obtain necessary and sufficient conditions for the design to be connected and balanced.

(8+2+4+6) = 20

3. (a) What do you understand by the recovery of interblock information?

(b) Is it always possible to obtain separately the intra- and inter-block estimates of a treatment contrast, in a balanced incomplete block design?

(c) How is the 'combined inter- and intra-block estimate' obtained?

(d) Is it always better than the intra-block estimate alone?

(e) What happens if the design is not a BIBD ?

(4x5) = 20

4. (a) What are cross-over designs?

(b) Explain the circumstances in which such designs are most appropriate.

(c) Assuming the absence of residual effects, give the structure of the analysis of variance for studying the effect of v treatments.

(d) How do you test for the absence of residual effects?

(4+4+8+4) = 20

5. A Balanced Incomplete Block Design (v, b r, k, λ) is modified to include a new treatment by increasing the block size to (k + 1) plots, so that the new treatment occurs once in each of the b blocks.

(a) Show how to analyse this experiment.

(b) Prove that the intra-block variance of the difference between effects of any two of the original set of v treatments is

$$\frac{2\,(k+1)\,\sigma^2}{v\lambda + r}$$

where σ^2 is the intra-block variance for blocks of (k+1) plots.

(c) Find the overall efficiency of this design.

(10+5+5) = 20

6. The actual design, before randomization, of a half replicate of a 2^6-factorial experiment involving six factors A, B, C, D, E and F, each at two levels and arranged in 4 blocks of 8 plots each, is as given below:

Block I.	Block II	Block III	Block IV
(1)	ac	ad	ae
ab	bc	bd	be
acde	de	ce	cd
bcde	abde	abce	abcd
adf	df	cf	cdef
bdf	abdf	abcf	abcdef
ef	acef	adef	af
abef	bcef	bdef	bf

Q.6 (contd.)

(a) Determine which of the interactions have been confounded.

(b) Write down the main effects and interactions which are aliases, while being estimated from this design.

(c) Assuming that three factor and higher order interactions to be negligible, write down the partition of the degrees of freedom in the table of analysis of variance against each ascribable source of variation.

(5+5+10) = 20

7. (a) What are 'weighing designs'?

(b) Discuss the fundamental assumptions in a weighing design and its analysis.

(c) Let $\beta_1, \beta_2, \ldots, \beta_p$ be the true weights of p objects. Let there be N weighings, and define a matrix $X = (x_{i\alpha})$ such that

$x_{i\alpha} = +1$, if the i-th object is included in the α-th weighing in the right pan,

$= -1$, if the i-th object is included in the α-th weighing in the left pan,

$= 0$, if the i-th object is not included in the α-th weighing.

Show that all the objects would be weighted with maximum precision when $XX' = N I_p$ where I_p is the identity matrix of order p.

(d) Define a Hadamard matrix of order 2^k and show that it can always be constructed. How are those matrices useful in Weighing Designs?

(5+6+5+4) = 20

8. (a) Explain the use of orthogonal polynomials in obtaining the detailed analysis of an asymmetrical factorial design.

(b) Obtain a balanced group of sets for 4^2 factorial design in blocks of 4 units, where the two-factor interaction is confounded.

(c) Give the breakdown of various degrees of freedom in the analysis of variance of above design.

(4+8+8) = 20

9. Write short notes on any three of the following:

(i) Fundamental Principles of Design of Experiments.
(ii) Elimination of heterogeneity in several directions
(iii) Designs confounding the Main Effects
(iv) Response Surface Analysis and Rotatable designs
(v) 'Missing values' and 'mixed-up yields' techniques
(vi) Balancing in symmetrical factorial designs.

Group B: Sample Surveys (Theoretical Aspects) -I.

(Answer any four questions from this group)

1. (i) Explain briefly the circumstances under which you would prefer a complete enumeration to a sample survey.

(ii) Discuss clearly the relative merits of (a) probability sampling, (b) haphazard sampling and (c) purposive or judgement sampling.

(10+15) = 25

2. (i) In the case of sampling n units from a finite populatio of N units with equal probability and without replacement, suggest an unbiased estimator of the population mean. Derive it sampling variance and obtain an unbiased variance estimator.

(ii) Suggest an operational procedure for selecting with the help of a table of random numbers, 10 units with equal probabil and without replacement from a population of 498 units.

(iii) Suppose there is an unknown number (say N) of similar objects in a box serially numbered from 1 to N. For estimating the value of N, a sample of n objects is picked up at random after thorough shuffling. Derive an unbiased estimator of N, using the serial numbers on the selected objects and obtain its variance and an unbiased variance estimator.

(10+5+10) = 25

3. (i) Suppose n units are selected from a population of N units with probability proportional to a given measure of size x and with replacement. Suggest an unbiased estimator of the population total and derive its sampling variance.

(ii) Compare the efficiency of the estimator in (i) with that in the case of sampling n units with simple random sampling and with replacement, (a) when the cost is proportional to the numb of units in the sample, (b) when the cost is proportional to the expected total size (x), of the sample units.

(iii) Describe briefly two operational procedures available f selecting units with probability proportional to size.

(8+10+7) = 25

4. (i) What is meant by multi-stage sampling and under what cir cumstances would you use this procedure in preference to direct sampling of units (i.e. uni-stage sampling).

(ii) For estimating the total number of persons (Y), a sample of n villages is selected from the N villages in the region wit probability proportional to the current number of households (M_i) and with replacement. From each selected village again m households are sampled circular systematically. The number of person in each of the nm sample households is ascertained by enquiry.

Suggest an unbiased estimator of Y and obtain an unbiased variance estimator for it.

(10+15)=25

5. (i) Explain clearly the regression method of estimation poi ting out how it differs from the ratio method of estimation.

contd.

Q.5.(contd.).

(ii) Suggest the regression estimator for the population mean ($\bar{Y}$) in the case of sampling n units with equal probability and without replacement, when information on an auxiliary variable x is available for all the N units in the population. Obtain its exact bias and an approximation to its mean square error.

(iii) To estimate the total yield of guava in a district, a two-phase sampling design was adopted. First a sample of n_1 villages was selected with probability proportional to area under orchards (as obtained for the previous year) and with replacement. For each of these n_1 villages, the current total number of orchards (a) and the current number of guava trees, (b) were obtained. A sub-sample of n_2 villages is selected at random from the initial sample of n_1 villages and from each of these villages, m orchards were selected with equal probability and without replacement for observing the yield of guava crop. Let y_{ij} denote the yield of the j-th sample orchard in the i-th sample village ($j = 1, 2, \ldots, m$; $i = 1, 2, \ldots, n_2$). Suggest a suitable regression estimator for the total yield of guava crop in that district.

(7+10+8)= 25

6. Write brief notes on any four of the following:

(i) balanced systematic sampling;

(ii) cluster sampling;

(iii) quota sampling;

(iv) self-weighting design;

(v) non-sampling errors; and

(vi) controlled selection. (25)

Group F: Techniques of Computation

Answer any five questions from this group.

1. The value of the polynomial

$$P = ax^2 + bx + c$$

for a given x is calculated by the following procedure

(i) $A = a.x$

(ii) $B = A+b$

(iii) $D = B.x$

(iv) $P = D+c$

If x, a, b and c are given correct to 's' significant digits and the intermediate results A, B and D are rounded off to 's' significant digits, estimate the error in P.

(20)

2. The function $f(x)$ is tabulated for $x = x_0$, x_0+h, x_0+2h, x_0+3h, x_0+4h and x_0+5h.

(i) Obtain a quadrature formula for evaluating

$$\int_{x_0}^{x_0+5h} f(x)\, dx$$

as accurately as possible.

(ii) Give an estimate of the remainder.

(10+10)= 20

3. The r-th Legendre polynomial, denoted by $P_r(x)$, is given by:

$$P_r(x) = \frac{1}{2^r\, r!} \frac{d^r}{dx^r}\left[(x^2-1)^r\right]$$

Show that

(i) $P_r(1) = 1$

(ii) $P_r(x) \leq 1$ for $|x| \leq 1$

(iii) $\int_{-1}^{+1} P_r(x)\, P_3(x)\, dx = 0$ if $r \neq s$

(iv) $P_{r+1}(x) = \frac{2r+1}{r+1} \cdot x \cdot P_r(x) - \frac{r}{r+1} \cdot P_{r-1}(x)$.

(3+3+6+8)= 20

4. (i) By applying the Newton-Raphson procedure to $f(x) = 1 - \frac{1}{x}$ obtain the recurrence formula

$$z_{k+1} = z_k\,(2 - Nz_k)$$

for the iterative determination of the reciprocal of N without effecting division.

(ii) Show that, if ϵ_k denotes the error in z_k, then

$$\epsilon_{k+1} \approx N\epsilon_k^2 \quad \text{when} \quad z_k \approx \frac{1}{N}$$

(iii) Obtain a third order iterative scheme for the determination of the reciprocal without effecting division.

(6+8+6) = 20

contd.

5. The sequence of column vectors $x^{(k)}$ generated by the recurrence formula,

$$x^{(k)} = (I - \alpha A)\, x^{(k-1)} + \alpha b ,$$

where $A = ((a_{ij}))$ is nxn symmetric matrix

I is the identity matrix

b is a column vector

α is a positive real number

converges to the solution of $A x = b$, if

$$\alpha < \frac{2}{\max_{i} \sum_{j=1}^{n} |a_{ij}|} \qquad (20)$$

6. The nxn matrix $A = ((a_{ij}))$ is such that

$$a_{ij} = 0 \quad \text{for } j > i+1 \text{ and } i = 1,2,\ldots, n-1$$

$$\text{and for } j < i-1 \text{ and } i = 2,3,\ldots, n$$

$$= \text{non-zero for other } i \text{ and } j$$

and $a_{i,i+1} \cdot a_{i,i-1} > 0$ for $i = 2, 3, \ldots, n-1$.

Show that all the eigen values of A are real. (20)

7. If $p_0, p_1, \ldots, p_n$ are all positive, show that the roots α_i, $i = 1, 2, \ldots, n$, of the equation

$$p_0 z^n + p_1 z^{n-1} + \ldots + p_n = 0$$

satisfy the inequality

$\alpha \leq |\alpha_i| \leq \beta$ where α and β are the maximum and minimum of the numbers

$$\frac{p_1}{p_0}, \frac{p_2}{p_1}, \ldots \frac{p_n}{p_{n-1}}. \qquad (20)$$

8. Show that the error in the following iterative scheme

$$y_{n+1} = y_{n-1} + 2h\, y'_n$$

for the numerical solution of the differential equation

$$y' = f(x, y); \; y'_0 = f(x_0, y_0)$$

grows exponentially with n.

(20)

GROUP G : STATISTICAL INFERENCE (General Theory) - I

(Answer any five questions from this group)

1.(a) $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal population $N(0, \sigma^2)$. Write $\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$. Examine rigourously whether each of the statistics, $\sum_{1}^{n} X_k^2$ and $\sum_{1}^{n} (X_k - \bar{X})^2$ is

(i) sufficient for σ,

and (ii) a complete sufficient statistic for σ

(b) If T_n represents the total number of successes in n independent Bernoulli trials with unknown probability p of success, then

EITHER

find an unbiassed estimator p^2, of the form $a T_n^2 + b T_n + C$,

OR

prove that there exists no unbiassed estimator for $\frac{1}{p}$.

(7+5+8) = 2

2.(a) Prove that under certain regularity conditions (to be stated), the maximum likelihood estimator is asymptotically normal and is consistent.

(b) Let $X_{i,j}$, $i = 1, 2$; $j = 1, 2, \ldots, n$ denote the j-th observation from the i-th population. The two populations follow normal distributions with unknown means μ_1, μ_2 and unknown but common variance σ^2.

(i) Find the maximum likelihood estimator of σ^2

(ii) Examine whether it is unbiassed and/or consistent.

(8+4+8) = 2

3.(a) Let H be the simple hypothesis that a random variable X has probability density function $f(x)$ and A be the simple alternative that X has probability density function $g(x)$. Let α, β be the sizes of the errors of type I and type II. Proceeding as in the proof of the Neyman-Pearson Lemma, prove that the test-procedure (based on n observations) which rejects H if $g(x) \ldots g(x_n) \geq 2\, f(x_1) \ldots f(x_n)$ and accepts H otherwise, is the one which minimizes the value of $\frac{2}{3}\alpha + \frac{1}{3}\beta$.

(b) Find this minimum value if $P(X = 1) = \theta$, $P(X = 0) = 1-\theta$, $H : \theta = \frac{1}{2}$; $A : \theta = \frac{3}{4}$ and $n = 7$.

(15+5) = 2

4. (a) Show that uniformly most powerful tests are unbiased.

(b) Prove that no uniformly most powerful test exists for testing $\theta = 0$ against $\theta \neq 0$, in a normal population $N(\theta, 1)$.

(c) Let $X_1, \ldots, X_n$ be a random sample from the uniform distribution on $(0, \theta)$. To test H: $\theta = 1$ against A: $\theta \neq 1$, the following test procedure is suggested: Reject H if $\max(X_1, \ldots, X_m) > 1$ or if $\max(X_1, \ldots, X_n) \leq \sqrt[n]{\alpha}$; α is preassigned, $0 < \alpha < 1$.

Examine the limit, as $n \to \infty$, of the power function of this test procedure.

(4+8+8) = 20

5. (a) Show that if a confidence interval for a parameter is obtained by 'inverting' a family of uniformly most powerful unbiased tests, each test being of level α, then the interval is the shortest unbiased interval of confidence coefficient $(1 - \alpha)$.

(b) $x_{(1)}$ is the minimum of a random sample of size n from a population with frequency function $f(x) = \frac{1}{\sigma} e^{-(x-\theta)}$, $x \geq \theta$. Find the confidence coefficient of the interval

$\left\{ x_{(1)} + \frac{1}{n} \log \alpha \, , \; x_{(1)} \right\}$ for the parameter θ, where α is a fixed number such that $0 < \alpha < 1$.

(11+9) = 20

6. (a) Define the terms: (i) admissible decision functions, (ii) minimal complete class of decision functions.

(b) Prove that a necessary and sufficient condition for the existence of a minimal complete class of decision functions, is that the class of admissible decision functions be complete.

(c) Suppose $P(X = 1) = \theta$, $P(X = 0) = 1 - \theta$ and θ may have the value $\frac{1}{4}$ or $\frac{1}{2}$. Let the decision space D consist of two elements a_1, a_2. Let the loss function be:

	a_1	a_2
$\theta = \frac{1}{4}$	1	4
$\theta = \frac{1}{2}$	3	2

Find the decision which is the minimax solution.

(4+8+8) = 20

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper IX: Subjects of Specialisation - II (Theoretical)

Time: 4 hours Full marks: 100

(a) Candidates are required to answer questions from that group only for which they have registered their options.

(b) Figures in the margin indicate full marks.

Group A: Economic Statistics - (Indian Economics and Economics of Planning)

(Answer any three questions from each section)

Section-I - Indian Economics.

1. Examine the present state of food statistics in India and indicate its pertinence for the formulation of the country's food policy. (16)

2. Give your understanding of the problem of increasing India's agricultural production. Do you think that this problem can be remedied through suitable technological changes without further land reforms? (16)

3. Analyse the changes in the sectoral distribution of India's national output in the course of the three five year plans. (16)

4. Examine how far does the present tax structure in India fit in with the requirements of rapid economic development. (16)

5. Discuss the reasons for the failure of the devaluation of the rupee to bring about, so far, any significant improvement in India's balance of payments position. (16)

6. Do you think that the proposals for social control of private commercial banks in India provide a sound alternative to their nationalisation? Give reasons in support of your view. (16)

Section II - Economics of Planning

7. What are the main features of Harrod's model? Analyse, in the light of this model, the phenomenon that if an economy increases its production too rapidly, it will have produced too little, whereas if production is not increased rapidly enough, the economy will have produced too much. (16)

8. Discuss in reference to the Mahalanobis model, how the higher priority to investment in capital-goods sector in India's Second Five Year Plan was based on rational considerations of long-run economic growth. (16)

9. Analyse how the static input-output system can fit into a linear programming model and indicate the problems of planning to which such a model is relevant. (16)

10. Review your understanding of any two of the following:

(a) a low level equilibrium trap
(b) shadow prices
(c) Hawkins-Simon conditions of viable production
(d) disguished unemployment

(8+8) = (16

11. Critically examine the employment policies followed in India's Second and Third Five Year Plans.

(16

12. Give an appraisal of India's industrial growth during the Second and Third Five Year Plans.

(16

Neatness (Sections I and II) (4

Group B: TECHNO-COMMERCIAL STATISTICS - (II)

Time: 4 hours. Full marks: 100. Pass marks: 50%.of t

Section- I (Alternative): ELEMENTS OF BOOK KEEPING & ACCOUNTS Total - 70 ma

Section- I : OPERATIONS RESEARCH - Total - 70 marks

Section-II : STATISTICAL METHODS IN BUSINESS - Total - 30 m

Section-I: (Alternative) - Elements of Book-keeping & Accounts

N.B.: (a) Use separate answer book for this section.
(b) Answer question No.6 and any other three questions from this group.

1.(a) Discuss the principles of Double Entry accounting with illustrations. Classify Accounts.

(b) Define and illustrate posting, casting and balancing processes in accounting.

(6+3)+(2x3) = (15

2. Pass necessary adjustment entries as on 31.12.67 in the following cases:

(a) Outstanding salary Rs.1,000

(b) Insurance premium Rs.1,800 paid on 1.9.67 for one year

(c) Rs.300 rent received on 1.12.67 for November and December, 1967 and for January, 1968.

(d) Interest on 6%. Investments Rs.10,000, is receivable on 31.3.68. It was purchased on 1.4.67

(e) Deferred advertisement expenditure Rs.4,000 has been carried forward from last year. The amount is to be written off over four years.

(3x5) = (15

3.(a) Distinguish between Journal and Ledger

(b) Name the sub-divisions of Journal with reasons for such sub-division

(c) Is Cash Book a Journal or a Ledger?

(6+5+4)= (15

4. Sri Jadav started a business with Rs.15,000 in cash, Rs.2,000 worth furniture and Rs.3,000 worth of stock, as on 1.4.67. He withdrew Rs.6,000 on 30.9.67, but on 1.12.67 he deposited Rs.11,000. The first account is closed on 31.12.67.

Interest at the rate 6°/. per annum is considered.

Profit for the period before considering interest on capital, amounted to Rs.8,000.

Make out Capital Account, showing calculations in details. Calculate interest in terms of nearest rupee. (15)

5.(a) What are the different types of errors that one usually finds when checking the accounts of a business concern. Give illustrations.

(b) What are the errors which are reflected by disagreement of Trial Balance?

(c) Name the types of errors with illustrations which are not disclosed by Trial Balance.

(d) Why is it that Closing Stock is not usually shown in Trial Balance? Pass Journal entry for bringing the Closing Stock of Rs.50,000 in the Balance Sheet.

(6+2+2+3+2) = (15)

6. From the following Trial Balance and notes make out Trading and Profit and Loss Account for the year ended 31st March, 1968 and a Balance Sheet as on that date.

	Dr. Rs.	Cr. Rs.
Drawings and Capital	5,000	50,000
Stock and Finished Goods (1.4.67)	12,500	
Stock and Materials (1.4.67)	5,000	
Purchases and Sales	30,000	91,000
Manufacturing Wages	10,800	
Salaries	4,400	
Returns - Inwards and Outwards	2,500	1,800
Discount Allowed and Received	800	500
Manufacturing Expenses	3,500	
Buildings	30,000	
Plant and Machinery	20,000	
Debtors and Creditors	40,000	20,000
Bad Debt & Reserve for Bad Debt	1,000	2,500
Rent and Rates	3,600	
Cash and Bank Overdraft	500	8,730
Bank Interest and Charges	850	
Insurance	900	
Postage & Telephone	1,180	
Furniture & Fixtures	2,000	
	1,74,530	1,74,530

Notes:

(1) Stock on 31.3.68: Finished Goods Rs.15,000; Raw Materials Rs.7,000.
(2) Rs.300 Salaries paid relate to 1966-67 whereas Rs.400 salaries for 1967-68 are still outstanding.
(3) The Insurance Policy is for one year and it expires on 30.6.68.
(4) Provide depreciation at the rate 5°/. on Buildings and 10°/. on Plant & Machinery and Furniture & Fittings. Two-third of Furniture & Fittings are for Office purposes.
(5) Provision for Bad Debt is to be maintained at 5°/. of Sundry Debtors. (25)

contd.

Group B: TECHNO-COMMERCIAL STATISTICS -(II) contd.

Section I: Operations Research (70)

Section II: Statistical Methods in Business (30)

SECTION I (OPERATIONS RESEARCH)

(i) Use a separate answer book for this section
(ii) Attempt any four questions from this section
(iii) All questions carry equal marks

1.(a) Use the Simplex method to maximize

$$z = 10 x_1 + 3 x_2 + 6 x_3 + 5 x_4$$

subject to the constraints

$$x_1 + 2x_2 + x_4 \leq 6 ;$$

$$3x_1 + 2x_3 \leq 5 ;$$

$$x_2 + 4x_3 + 5x_4 \leq 3 ;$$

$$x_1 \geq 0, \; x_2 \geq 0, \; x_3 \geq 0, \; x_4 \geq 0.$$

(b) A bicycle manufacturer makes two models, a sports cycle and a racer. In order to make a sports model, 6 man-hours are needed, while a racing model requires 10 man-hours. The manufacturer can employ no more than 15 men and these men work 8 hours per day for 5 days each week. The cost of materials amounts to Rs.50 per cycle and the manufacturer's weekly quota of such materials may not exceed Rs.4000. The firm has a contract to supply at least 30 'sports models' and 20 'racers' per week. How many cycles of each type should be made in order to obtain the maximum possible profit, if the profit on each sports cycle is Rs.10/- and on each racing model is Rs.30?

2.(a) Solve the transportation problem for which the costs, origin availabilities (a_i), and destinations requirements (b_j), are given in the following table, by using the North-West corner rule, to obtain an initial feasible solution: -

	D_1	D_2	D_3	D_4	D_5	(a_i)
O_1	4	3	4	5	5	100
O_2	4	4	4	3	1	100
O_3	5	5	4	3	4	120
(b_j)	40	40	80	80	80	320

(b) One car is available at each of the stations 1,2,3,4,5,6 and one car is required at each of the stations 7,8,9,10,11,12. The distances between the various stations are given in the matrix below. (see page 5).

Table: Matrix of distances

To Stations	7	8	9	10	11	12
1	41	72	39	52	25	51
2	22	29	49	65	81	50
3	27	39	60	51	32	32
4	45	50	48	52	37	43
5	29	40	39	26	30	33
6	82	40	40	60	51	30

How should the cars be despatched so as to minimize the total mileage covered?

3.(a) The cost of maintenance of a machine is given as a function increasing with time and its scrap value is constant. Show that the average annual cost will be minimized by replacing the machine when the average cost to date becomes equal to the current maintenance cost.

(b) The cost of a machine is Rs.6,100 and its scrap value is only Rs.100. The maintenance costs are found from experience to be as follows:

Year	1	2	3	4	5	6	7	8
Maintenance cost (in Rs.)	100	250	400	600	900	1250	1600	2000

When should the machine be replaced?

4.(a) At time zero, all items in a certain system are new. Each item has a probability p of failing immediately before the end of the first month of life and a probability $q = 1-p$ of failing immediately before the end of the second month (i.e. all items fail by the end of the second month). If all items are replaced as they fail, show that the expected number of failing f x, at the end of a month x, is given by

$$f(x) = \frac{N}{1+q}\left[1 - (-q)^{x+1}\right],$$

when N is the number of items in the system.

(b) If the cost per item of individual replacement is C_1 and the cost per item of group replacement is C_2; find the conditions under which a group replacement policy at the end of each month, is most profitable.

5.(a) Discuss the various costs involved in inventory problems.

(b) Discuss the economic lot size inventory model.

(c) Let the cost of owning a wagon be one unit per day and let the cost of hiring the wagon be k units per day ($k > 1$). The probability is p_n ($n = 1,2,\ldots$), that n will be required in a day.

Show how you would find the optimum number of wagons to be owned so that the expected cost of owning and hiring wagons is minimum.

contd.

6.(a). State some of the important distributions of arrival intervals and service times.

(b) Obtain the steady state probabilities for an unlimited queue served in order of arrival and a single service channel, the probabilities of arrival and of completion of service being constant. Find also:

(i) avarage number of persons in the queue,

(ii) probability distributions for the waiting time of a unit,

(iii) average waiting time for a unit,

(iv) probability of a waiting time greater than t.

7.(a) State the principle of optimality in dynamic programming and give the dynamic programming formulation of one inventory and one replacement problem.

(b) In the case of 2 channels, Poisson arrivals and exponential service, show that the expected number in the system, in the usual notation, is

$$\frac{4\lambda\mu}{4\mu^2 - \lambda^2}$$

SECTION II: STATISTICAL METHODS IN BUSINESS - (30 Marks).

(i) Use a separate answer book for this section.
(ii) Answer any two questions from this section.
(iii) All questions carry equal marks.

1.(a) A number n of firms are in the market selling tooth pastes. All of them advertise their products. Due to its own advertisement, the sales of a firm increase at a rate proportional to its sales at that time, but due to the advertisement by each of its rivals, its sales decrease at a rate proportional to their sales at that time. Set up a mathematical model and discuss its solution for $n = 2$.

(b) If a firm does not advertise its goods, its sales falls at a rate proportional to the sales at any time. It advertises at regular intervals and every time it advertises its rates go up by an amount 'a.' At what interval should it advertise so that the sales in every time-interval are the same? At what time intervals should it advertise so that the sales tend to $\frac{3}{2}$ of its original value?

2.(a) How would you proceed to estimate the demand for a cooking gas in a given town ?

(b) What do you understand by time series and its trend? How will time series analysis help you to forecast demand for a certain commodity?

3.(a) Explain the use of statistical methods in (i) job evaluation and personnel selection, (ii) effects of bonus and incentive schemes, (iii) accountancy and auditing.

(b) A firm starts manufacturing a new product. It wants to go on adjusting the amount of production every month in the light of the sales information received in the earlier months so as to maximize its profits over the year. Develop a suitable mathematical model for this purpose.

Group C: BIOMETRIC METHODS - (II)
Statistical Methods in Genetics & Bio-assays
(No candidate available)

Group D: DESIGN AND ANALYSIS OF EXPERIMENTS-(II)
(COMBINATORIAL ASPECTS).

(a) Answer any five questions from this group.
(b) Figures in the margin indicate full marks.

1. Define a group.

Demonstrate the falsity or correctness of the statement that the following are groups under multiplication:-

(i) the set I of all integers
(ii) the set C of all complex numbers
(iii) the set S of all real numbers of the form $x+y\sqrt{2}$ where x, y, are rational not both zero;
(iv) the set U of the third roots of unity.

(4+4+4+4) = 20

2. Give the Addition and Multiplication tables for GF(4). Hence:

(i) find 3 mutually orthogonal latin squares of order 4
(ii) construct the balanced incomplete block design: $v = b = 21$, $r = k = 5$, $\lambda = 1$.

(8+6+6) = 20

3. Denoting by N(v) the maximum number of mutually orthogonal latin squares of order v, establish

(i) $N(21) \geq 3$

(ii) $N(18) \geq 2$

(10+10) = 20

4. State the two 'module' theorems of Bose, for the construction of balanced incomplete block designs (BIBD). Illustrate the uses of these theorems by constructing.

(i) the BIBD : $b = 35$, $v = 15$, $k = 3$, $r = 7$, $\lambda = 1$

(ii) the BIBD: $b = (3t+1)(4t+1)$, $v = 12t+4$, $r = 4t+1$, $k = 4$, $\lambda = 1$ where $4t+1$ is a power of a prime.

(4+6+10) = 20

5. Explain, with illustrations, how PG (N, s) and EG(N, s) may be employed for the construction of partially balanced incomplete block designs.

Let N_i be the incidence matrix of the BIBD: $(v_i, b_i, r_i, k_i, \lambda_i)$, $i = 1, 2$. Show that the design whose incidence matrix is the Kronecker product $N_1 \times N_2$ is a PBIBD with at most three associate classes and obtain its parameters.

(10+10) = 20

6. Explain any method known to you for constructing a (s^n, s^{n-t}) design such that no main effect, first, second, ..., (d-1)th order interaction, is confounded.

Establish that in $_k s^n$ design where s is a power of a prime and the block size is s^k, the maximum number of factors which can be accommodated so that no main effect or two factor interactions are confounded, is $(s^k-1)/(s-1)$.

(10+10) = 20

7. (a) Elucidate the statement:- 'Latin Squares and Graeco-Latin Squares may be considered as fractional replications of factorial experiments'.

(b) Bring out the relationship between confounding and fractional replication.

(c) Construct a 4^5 design in $\frac{1}{4}$ replicate, with confounding to reduce the block size to 16 plots in which only 3 degrees of freedom from one first order interaction are confounded.

(5+5+10) = 20

8. Give a critical review of all the methods known to you for the study of response surfaces.

20

Group E: SAMPLE SURVEYS (II) - ORGANISATIONAL ASPECTS

(Answer any five questions from this group).

1. What do you understand by 'sampling frame'? What are the defects to which a sampling frame is likely to be subject and the possible reasons for such defects? What are the possible steps that you can think of to rectify the defects?

(4 + 8 + 8) = 20

2. In analysing the data from a large scale sample survey, what are the different methods of checking systems that you can think of at different stages of data processing, to ensure accuracy? Discuss their relative merits and demerits.

(12 + 8) = 20

3. Suppose you are asked to carry out a sample survey to study the different characteristics of 'operational agricultural holdings' in your State. What are the various problems which you should consider while planning the survey? If such a survey was conducted in the past, how will you make use of the results of that survey, in improving the sampling design?

(12 + 8) = 20

4. Explain the meaning, indicate the importance, and state the inter-connection (if any) of:

(i) pilot surveys,
(ii) variance function,
(iii) cost function,
and (iv) optimum allocation,

as understood, in the context of large scale sample surveys.

(4 x 5) = 20

5. Indicate the methods of checking the accuracy of final tables to be presented in a report. What, in your opinion, should be the nature of information to be presented in a report of a sample survey? Give an outline of the headings of sections which you would include in the report.

(5+10+5) = 20

6. What considerations should weigh with the officer-in-charge of a sample survey in deciding the type of investigators needed for each of the following types of surveys to be carried out in India:

(i) a survey for studies on family planning problems (attitude survey)

(ii) a survey for assessing the High Yielding (crop) programme

(iii) a survey for studying the food consumption pattern in rural areas.

What procedures do you think should be adopt for recruiting and training scheme?

(5+5+5+5) = 20

7. Write brief, but clear notes on any three of the following:-

(a) Non-response in sample surveys
(b) Schedules and questionnaires
(c) Inter-penetrating net work of sub-samples
(d) Data processing by punch-card equipments

20

Group F: TECHNIQUES OF COMPUTATION (II)

Time : 5 Hours Full marks : 100 Pass marks : 50

(a) Answer any five questions from this group.
(b) Figures in the margin indicate full marks.
(c) Use of Calculating Machines is permitted.

1. Tabulate the function $f(x)$ for $x = 1.00\ (0.02)\ 1.20$ making use of values of the function given below.

x	f(x)
0.9	0.78333
1.1	0.89121
1.3	0.96356

(20)

2. Compute the value of

$$\int_0^1 \frac{dx}{1+x^2}$$

using 3-point Legendre-Gauss formula and also estimate the error in the computed value. The weights and abscissas for (-1, 1) is given below.

Abscissas	Weights
0	8/9
± 0.774597	5/9

/(C+10):

3. Evalue all the roots (real and complex) of the equation

$$21x^3 - 6x^2 + 35x - 10 = 0$$

(20)

4. The symmetric matrix A and its inverse are given below:

$$A = \begin{bmatrix} 15.129 & 23.860 & 1.793 & 0.998 \\ & 54.756 & 3.633 & 3.511 \\ & & 18.225 & 21.122 \\ & & & 60.516 \end{bmatrix}$$

$$A^{-1} = \begin{bmatrix} 0.2121872 & -0.0922374 & -0.0077835 & 0.0045688 \\ & 0.0586077 & -0.0007232 & -0.0016267 \\ & & 0.0933391 & -0.0324080 \\ & & & 0.0278550 \end{bmatrix}$$

The matrix B is obtained by replacing the second row of A by the row vector,

$$\begin{bmatrix} 23.760 & 54.856 & 3.533 & 3.411 \end{bmatrix}$$

Compute the inverse of B without actually inverting B.

[Hint: $B = A+E$ and $B^{-1} = A^{-1}(I+EA^{-1})^{-1}$ where E has one non-null row and others are null]

(20)

5. Compute the inverse of C,

$$C = \left[\begin{array}{c:c} A & x \\ \hdashline x' & \alpha \end{array}\right]$$

where A is the matrix given in Q.4

$$x' = \begin{bmatrix} 1.0 & 1.0 & 1.0 & 1.0 \end{bmatrix}$$

and $\alpha = 1.0$

(20)

6. For the matrix A given in Q.4 compute the latent root with minimum absolute value and the latent vector corresponding to it.

(20)

7. Solve numerically the differential equation

$$dy / dx = 5y / (1+x);$$

$$x = 0 \quad , \quad y = 1 ;$$

for $x = .1(.1)\ 1.0$ and compare with the exact values of the function.

(15+5) = 20

Group C: STATISTICAL INFERENCE (II)

(Special Topics)

(Answer any five questions from this group)

1. (a) What is a sequential test? What is a sequential probability ratio test (SPRT)?

(b) Obtain the SPRT of (approximate) required strength (α, β), for a simple hypothesis $\theta = \theta_0$ against a simple alternative $\theta = \theta_1$. State clearly the underlying assumptions. How would you find the OC and ASN functions of this test?

(c) Show that the SPRT terminates with probability one, under some conditions (to be clearly stated).

(4+10+6) = 20

2. (a) $x_1, x_2, \ldots, x_m$ is a random sample from a continuous distribution F and $y_1, y_2, \ldots, y_n$ is an independent random sample from a continuous distribution G. Define the Wilcoxon and Mann-Whitney statistics for testing the hypothesis $H_0: F = G$ and show that these two tests are equivalent and distribution-free.

(b) Obtain the mean and variance of the Mann-Whitney statistic in the null-case. What is its asymptotic distribution?

(10+10) = 20

3. (a) Define a U-statistic.

(b) State and prove its optimality as an estimator, stating clearly the underlying conditions and any side-results that you might use.

(c) Discuss briefly the use of Kendall's rank-correlation statistic as a measure of association, between two variables.

(2+10+8) = 20

4. (a) If a column-vector X is distributed normally with mean μ and dispersion matrix Σ, obtain the distribution of $(X - \mu)' \Sigma^{-1} (X - \mu)$ and, hence, point out how you would test the hypothesis $H_0: \mu = \mu_0$ if Σ is known.

(b) Show that the likelihood-ratio test of H_0, if Σ is unknown, is based on Hotelling's T^2-statistic.

(c) Derive the null-distribution of T^2.

(6+6+8) = 20

5. (a) Obtain the likelihood-ratio test for the hypothesis of equality of means of k p-variate normal populations with a common dispersion matrix, and derive the moments of the test criterion in the null case.

(b) Give explicit test procedures for the cases

(i) p = 1, 2 with any k,

and (ii) k = 2, 3 with any p.

contd.

Q.5 (contd.)

(c) Write down the analysis of dispersion and indicate briefly how you would carry out the test in the general case.

(12+4+4) = 20

6.(a) What are principal components?

(b) Give a suitable computational procedure to determine them in practice.

(c) Discuss briefly their use in Factor Analysis.

(8+6+6) = 20

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - May 1968

Paper X - Subjects of Specialisation - III (Practical)

Time: 5 hours Full marks: 100

(i) Candidates are required to answer questions from that group only for which they have registered their options.

(ii) Figures in the margin indicate full marks.

(iii) Use of calculating machine is permitted.

GROUP A : ECONOMIC STATISTICS
Paper III - P r a c t i c a l
(No candidate available)

GROUP B : TECHNO-COMMERCIAL STATISTICS (Paper III)

Section I : Statistical Quality Control - 50 marks

Section II : Operations Research - 30 marks

Or in the alternative

Elements of Book-keeping and Accountancy - 30 marks

Section III: Statistical Methods in Business - 20 marks

(Separate answer books are to be used for each of the Sections of B)

Section I : Statistical Quality Control

(Answer question No.1 and any other two from this Section)

1. (a) The cathode warm-up time in seconds was determined for three different tube types using eight observations on each type of tube. The order of experimentation was completely randomized. The results were :

Tube Type	warm-up time in seconds (8 cases for each type)
A	19, 20, 23, 20, 26, 18, 18, 35
B	20, 40, 20, 24, 32, 22, 27, 18
C	16, 19, 15, 17, 18, 19, 26, 18

Obtain the analysis of variance of these data and test the hypothesis that all the three tube types require the same average warm-up time.

(b) For the above case, set up othogonal contrasts between the tube types and test your contrasts for significance.

(12+6) = 18

Please turn over

2. In a research Centre at a University, five electrode shapes A, B, C, D and E were studied. For this experiment five holes were cut in [illegible] work-pieces. The cutting was accomplished by an electric discharge between the electrode and the material being cut. Besides, the order of electrodes was arranged so that only one electrode shape was used in the same position, on each of five work-pieces.

The results so obtained with regard to the variable "RC-hardness" appear below :

	Position of 'hole' on the work-piece (strip)				
Stripe	1	2	3	4	5
I	A(64)	B(61)	C(62)	D(62)	E(62)
II	B(62)	C(62)	D(63)	E(62)	A(63)
III	C(61)	D(62)	E(63)	A(63)	B(63)
IV	D(63)	E(64)	A(63)	B(63)	C(63)
V	E(62)	A(61)	B(63)	C(63)	D(62)

Analyse these data and test for an electrode effect, position effect and stripe effect on RC-hardness. (16)

3. The following are the $\bar{x}$ and R values, for 20 sub-groups of five readings. The specification for this product characteristic is 0.40317 ± 0.0010. The values given are the last two figures of the discussion reading. (i.e. 31.6 means the number 0.40316)

Sub-Group No.	$\bar{x}$	R	Sub-Group No.	$\bar{x}$	R
1	34.0	4	11	35.8	4
2	31.6	4	12	35.8	4
3	30.8	2	13	34.0	14
4	33.0	3	14	35.0	4
5	35.0	5	15	33.8	7
6	32.2	2	16	31.6	5
7	33.0	5	17	33.0	5
8	32.6	13	18	33.2	3
9	33.8	19	19	31.8	9
10	35.8	6	20	35.6	6

(a) Draw suitable control charts to determine if the process is in control.

(b) Under suitable assumptions (to be stated) estimate the proportion of items that will fail to meet the specification.

(16)

4. The following Double Sampling plan is used :

A product in lots is submitted for inspection
$n_1 = 50$, $c_1 = 0$, $n_2 = 60$, $c_2 = 3$.

Assuming that the lot is very large compared to the sample sizes, determine the probability L(p) of acceptance if the submitted quality is p% defective.
Compute L(p) for 5 suitable values of p, and draw the OC curve. (16)
Use necessary approximation.

Please turn over

GROUP B : TECHNO-COMMERCIAL STATISTICS (Contd.)

Section II:Operations Research

(Use separate answer-book for this Section)

(Answer any two questions from this Section)

1. A baking company sells cake by the pound. It makes a profit of 1 rupee on every pound sold on the day it is baked. It disposes of all cakesnot sold on the date it is baked,at a loss of 10 paise per day pound. If the demand is known to be rectangular between 200 and 300 pounds,determine the optimal daily amount to be baked. (15)

2. Truck tyres which fail in service can cause expensive accidents. It is estimated that a failure in service results in an average cost of Rs. 1,000/-, exclusive of the cost of replacing the blown tyre. New tyres cost Rs.500/- each, and are subject to mortality as shown in the following table. If the measure of effectiveness for a replacement policy is the average cost per mile, and if tyres are to be replaced after a certain fixed mileage or on failure (whichever occurs earlier, determine the optimum replacement policy.

Table 1 : Truck-tyre mortality

mileage of tyre covered at failure	proportion of failure	mileage of tyre covered at failure	proportion of failure
- 10,000	0	17,001 - 18,000	0.060
10,001 - 11,000	0.010	18,001 - 19,000	0.090
11,001 - 12,000	0.010	19,001 - 20,000	0.130
12,001 - 13,000	0.015	20,001 - 21,000	0.180
13,001 - 14,000	0.020	21,001 - 22,000	0.165
14,001 - 15,000	0.027	22,001 - 23,000	0.125
15,001 - 16,000	0.036	23,001 - 24,000	0.080
16,001 - 17,000	0.040	24,001 - 25,000	0.012

(Total of proportions adds up to 1.000)

(15)

3. A super-market has two sales girls at the counters. If the service time for each customer is exponential with mean 4 minutes, and if people can be assumed to arrive in a Poisson fashion at the counter, at the rate of 10 an hour -

(a) What is the probability of having to wait for service ?

(b) What is the expected percentage of idle time for each girl ?

(15)

Group B Section II (Alternative) - Elements of Book-keeping and Accountancy.

(Alternative to Operations Research)

Please turn over

GROUP B : TECHNO-COMMERCIAL STATISTICS (Contd.)

Section II (Alternative)

Elements of Book-keeping and Accountancy (Practical)

(a) Use a separate answer book for this section

(b) Answer any two questions from this section

1. (a) Write short notes on the following :-

(i) Sales Day Book

(ii) Journal Proper

(iii) Fixed Asset

(iv) Cross Profit

(b) Pass the necessary journal entries to record the following transactions :-

Cash realised from Debtors	Rs. 1,000
Withdrawn from Bank	" 1,200
Bought goods from S. Lahiri	" 400
Purchased Office Equipment	" 500
Paid to Creditors	" 300
Goods returned to P. Sanyal	" 100
Rent due to Landlord	" 200

2. (a) What is a Bill of Exchange and how does it differ from a Promissory Note ?

(b) What do you understand by the terms "Capital Expenditures" and "Revenue Receipts"?

(c) Give the rulings of a Petty Cash Book kept under Imprest syste and state the advantages of such a system to a trader. (1

3. The Profit and Loss Account for the year ended 31st December, 1966 of Shri G. Das, had shown a profit of Rs. 12,000/- prior to the carrying out of the following adjustments :-

(i) Depreciation on Machinery @ 10% per annum on Rs. 4,000/- was not charged.

(ii) Provision for Bad and Doubtful debts amounting to Rs. 250/- is to be made.

(iii) Stock-in-trade on 31.12.66 adjusted in the account was understated to the extent of Rs. 500/-.

(iv) Rent receivable on 31.12.66 amounting to Rs. 150/- was not recorded.

You are required to pass necessary journal entries to record the above adjustments, and prepare an adjusted Profit and Loss Account and close the account. (1

Please turn over

GROUP B : TECHNO-COMMERCIAL STATISTICS (Contd.)

Section III - Statistical Methods in Business (Practical)

(1) Use a separate answer book for this section

(2) Answer any two questions from this section

1. The manager of the cafeteria of a large manufacturing company claims that at least 75% of the employees who eat in the cafeteria would prefer to have small portions on the Rs.5/- "special plate" and a corresponding reduction in price to Rs. 3.50. What can the management of the company conclude, at a level of significance of 0.05, if in a random sample of 300 employees who eat at the cafeteria, 230 are in favour of the change? (10)

2. For the years 1952 to 1962 (both inclusive), the total annual consumption of cotton in the United States (in millions of bales) was 9.2, 9.3, 8.5, 9.1, 9.0, 8.4, 8.1, 9.0, 8.7, 8.5, and 8.7.

   (a) Find the least squares line, which would best describe the downward trend of cotton consumption over this period.

   (b) Rewrite the trend equation after changing the origin to June 1955, x units to 1 month and y to average monthly consumption (in millions of bales). (10)

3. An Opinion research organization wants to determine whether there is any relationship between the quality of interviewers' work and their scores on an 'introvert - extrovert' test of personality. Each interviewer is rated by his supervisor as being 'above average', 'average' and 'below average' on the basis of factors such as persistence, need for supervision, complaints from alleged respondents, neatness in completing schedules and so on. The rating results are as shown in the following table :

Test	Personality ranking		
	above average	average	below average
	(1)	(2)	(3)
Introvert	18	28	14
Average	37	63	30
Extrovert	15	29	16

Find the expected frequencies under the null hypothesis of 'no relationship'. Calculate X^2 (chi-square) and test the null hypothesis at a level of significance of 0.05.

What can one conclude about the effectiveness of this personality test, in predicting whether a person applying for a position with the Opinion research organisation, will turn out to be a good interviewer? (10)

4. A company has four territories open and four salesmen available for assignment. The territories are not equally rich in their sales potential. It is estimated that a typical salesman operating in each territory, would bring the following annual sales :

Territory I	... Rs.60,000	Territory III	... Rs.40,000
Territory II	... Rs.50,000	Territory IV	... Rs.30,000

The four salesmen are also considered to differ in ability. It is estimated that working under similar conditions, their yearly sales would be proportionately as follows :

Salesman A	... 6	Salesman C	... 5
Salesman B	... 5	Salesman D	... 4

If the criterion is maximum expected total sales, the intuitive answer is to assign the best salesman to the richest territory, and the next best salesman to the second richest territory, and so on. Verify this answer by the 'assign' method.

Group C : Biometric Methods
Paper III - Practical
No candidate available

Group D : DESIGN & ANALYSIS OF EXPERIMENTS

Paper III - Practical

(Answer any three questions from this group)

1. An experiment in $\frac{1}{2}$th replicate of a $2^{8.0}$ factorial experiment, with factors A, B, C, D, E, F, G, and K - each at two levels, and arranged in 4 blocks of 8 plots each, was carried out to study the effect of these factors on the yield of paddy crop. In the table below are given the lay-out plan and yield of paddy (in lbs. per plot), the treatments within a block being assigned at random to the 8 plots in the block :

Lay-out plan and yield of paddy

Block I		Block II		Block III		Block IV	
Treatment	Yield (per plot)	Treatment	Yield (per plot)	Treatment	Yield (per plot)	Treatment	Yield (per plot)
adg	26.99	aeg	31.34	bdefk	26.22	beefh	36.98
cdef	32.81	abcdf	20.85	acfk	29.90	abgk	27.77
aceh	33.26	adfgh	32.29	adefg	30.32	abcdo	32.24
bdh	33.13	ef	31.02	egh	32.95	adfh	34.29
abf	31.30	adeh	30.45	cd	31.73	bdfg	31.73
abcdefgh	35.43	bch	38.13	abcdgh	29.81	acefg	33.58
fgk	34.15	bdeg	30.45	abc	28.46	cdogh	31.60
bceg	35.57	abefgh	33.64	befg	28.91	1	35.82

(a) Identify the defining contrasts and the confounded interactions;

(b) Analyse the data and interpret the results of your analysis.

(8+24) = 32

2. Sixteen paddy varieties, numbered 1, 2, ..., 16, were treated for yield in a Double Lattice Design. The table below gives the layout plan (without randomization) and the yield figures in grams per plot corresponding to the variety numbers shown in parantheses.

Yield of paddy (in grams per plot)

REPLICATION I

BLOCKS				
1.	650(1)	670(2)	720(3)	680(4)
2.	685(5)	655(6)	670(7)	905(8)
3.	685(9)	680(10)	680(11)	560(12)
4.	725(13)	690(14)	735(15)	605(16)

REPLICATION II

5.	735(1)	690(5)	840(9)	805(13)
6.	650(2)	605(6)	820(10)	735(14)
7.	670(3)	675(7)	685(11)	745(15)
8.	910(4)	790(8)	825(12)	855(16)

Please turn over

(a) Analyse the data with and without recovery of inter-block information; a

(b) Determine the efficiency of the design, with the same experimental material. (26+6) = 32

3. A varietal trial was conducted at a Research Station for three consecutive years - the actual site of experimentation was however varied from year to year. The design adopted for the same was five randomized blocks of 6 plots each, there being independent randomization every year. The yields in lbs. per plot (of 1/20th of an acre) obtained from these experiments are as under

Years	BLOCKS	VARIETIES V_1	V_2	V_3	V_4	V_5	V_6
1950-51	I	31	33	26	30	31	31
	II	33	31	35	45	26	2
	III	31	31	30	44	25	31
	IV	30	41	30	28	31	23
	V	34	10	28	23	15	20
1951-52	I	33	23	24	24	13	8
	II	29	29	27	24	23	23
	III	28	21	38	24	16	18
	IV	37	34	38	33	28	30
	V	48	46	45	45	50	25
1952-53	I	30	23	34	25	20	13
	II	39	22	28	25	28	32
	III	56	43	43	31	49	17
	IV	38	45	36	35	32	20
	V	44	51	23	58	40	30

Analyse the data for each year separately. Also combine the results for different years and state your conclusions. (32)

4. Eight objects (a, b, c, d, e, f, g and h) are weighed 16 times in a chemical balance, ~~This done~~ by putting some objects on the left pan and some on the right pan, and balancing ~~in every case~~ against standard weights (grammes) placed on the ~~other~~ pans. The results are presented in the following table :-

Results of 16 Weighings of 8 objects

serial no. of weighing	objects placed on right pan	objects placed on left pan	standard weights (grammes) to balance right pan	standard weights (grammes) to balance left pan
1	abcdefgh	...	...	4.296
2	abcd	efgh	0.269	...
3	acog	bdfh	0.050	...
4	abef	cdgh	0.005	...
5	acfh	bdeg	0.090	...
6	abgh	cdef	0.312	...
7	adeh	bcfg	...	0.235
8	adfg	bceh	...	0.166
9	...	abcdefgh	4.521	...
10	efgh	abcd	...	0.303
11	bdfh	acog	...	0.127
12	cdgh	abef	0.023	...
13	bdeg	acfh	...	0.198
14	cdef	abgh	...	0.209
15	bcfg	adeh	0.104	...
16	bceh	adfg	0.238	...

GROUP D : DESIGN & ANALYSIS OF EXPERIMENTS (Contd.)

4.(contd.)

(a) Estimate the true weights of the objects;

(b) Find the standard errors of the estimates and estimate the error variance of a single weighing; and

(c) Examine whether the balance is biased. (18+12+2)=32

Neatness (4)

GROUP E : SAMPLE SURVEYS

Paper III: Practical

(Answer any three questions from this group)

1. Table below shows the number of inhabitants of 64 cities in a country in a particular census year. The cities are arranged into two strata, the first containing the population of the 16 largest cities and the second the population of the remaining 48 cities.

Number of inhabitants (000)				Number of inhabitants (000)			
Stratum I	Stratum II			Stratum I	Stratum II		
799	315	171	122	455	233	136	114
772	299	171	122	440	233	140	112
750	298	165	117	413	218	138	110
736	260	165	120	400	210	136	109
590	254	163	119	385	203	138	108
575	244	159	116	380	195	133	105
537	240	155	118	325	181	130	100
535	238	145	115	318	180	125	102

Suppose the number of inhabitants in all the 64 cities is to be estimated from a sample of 24 cities. Find the relative standard error of the estimated total for (i) an unstratified simple random sample, (ii) a stratified simple random sample with proportionate allocation, (iii) a stratified simple random sample with 12 units drawn from each stratum, the selection being without replacement in all the cases. Indicate, on the basis of the results obtained, which of the three methods you would prefer. (9+9+9+5)=32

2.(a) From the table above (Q.No.1) draw a sample of size 24 with probability proportional to size and with replacement. (A full description of the method of selection is required to be added).

(b) From the sample obtained as above estimate the total urban population of the country, and furnish an estimate of error for the estimated total. Have you any comments to offer on the estimate? (22+13)=35

3. It is felt that progress in State regarding establishing Seed Farms for multiplication and distribution of high yielding varieties has not been quite satisfactory, because of various difficulties like - (i) want of adequate supply of fertilisers and pesticides, (ii) want of irrigation facilities, (iii) want of storage space for grains, etc. In order to spot various difficulties, it is intended to survey a sample of District Agriculture Farms and Block Farms.

Draw a schedule or questionnaire, covering all the important aspects of the problem, and write a short note on the concepts and definitions used in the schedule or questionnaire. (24+8)=32

4. With regard to above enquiry (Q.No.3) it is required to submit a scheme to Government with an indication of the total budget and a note on the sampling procedure.

Draw up a scheme taking into account particularly the following features: i) The enquiry is proposed for three Administrative Divisions, into which the entire State is divided; ii) Time taken for canvassing a schedule may be taken as 4 man-days; iii) A sample of 100 blocks is to be surveyed in all; iv) Journey-time is to be taken into account; v) Provision should be made for an adequate number of supervisory staff.

[The few components of cost mentioned in the above list are merely suggestive but not exhaustive.] (32)

Neatness (4)

GROUP F : TECHNIQUES OF COMPUTATION

Paper III : P r a c t i c a l

(No candidate available)

GROUP G : STATISTICAL INFERENCE

Paper III : P r a c t i c a l

(Answer any four questions from this group)

1.(a) Let X be normally distributed with mean θ and variance one. You are to test H_0 ($\theta = 0$) against H_1 ($\theta = \pm 1$). The loss is simple, i.e., for a correct decision it is zero and for an incorrect decision it is one. Assuming that the least favourable distribution assigns equal probabilities $(1-\varepsilon_0)/2$ to $\theta = \pm 1$, and probability ε_0 to $\theta = 0$, where ε_0 is to be properly chosen, find the minimax decision function for this problem.

(b) In the above problem find the Bayes solution when $\varepsilon_0 = 1/3$. (16+8)=24

2. Let X have the Cauchy density

$$f(x|\theta) = \frac{1}{\pi} \cdot \frac{1}{1 + (x - \theta)^2}, \quad -\infty < x < \infty.$$

Find the most powerful test based on a single observation and of size .05, for $H_0(\theta = 0)$ against $H_1(\theta = 2)$. Plot its power function. (24)

3.(a) The following data represent the number of red blood cells (in millions per cubic millimetre) for nine men and eight women.

Men	5.02,	4.58,	5.57,	4.52,	4.81,	5.36,	4.27,	4.93,	5.15
Women	4.15,	4.56,	3.89,	4.40,	4.38,	4.20,	4.31,	4.73	

Applying in turn each of the following tests : the Sign test, Wilcoxon's U-test and Student's t-test, test whether there is any difference between the two sexes with respect to red blood cells. State the assumptions behind each test.

(b) The number of automobile accidents per week were for a period of ten weeks,

12, 8, 20, 2, 14, 10, 15, 6, 9, 4

Are these frequencies in accordance with the belief that the accident conditions were the same throughout the ten-week period?

(c) Test to see whether the two variables of classification in the following 2 x 3 contingency table are independent.

	c_1	c_2	c_3	Total
v_1	20	10	10	40
v_2	10	20	30	60
Total:	30	30	40	100

(The entries in this table are frequencies in the different classes)

(12+6+6)=24

GROUP C : STATISTICAL INFERENCE (Contd.)

4. Measurements were made on fifty specimens of flowers from each of two species of iris. Two measurements were taken on each flower, namely, petal length (x_1) and petal width (x_2). The means were (in cm)

Variate	Species 1	Species 2
x_1	4.260	1.462
x_2	1.326	0.246

The pooled sums of squares and products about means were

	x_1	x_2
x_1	12.2978	3.8794
x_2	3.8794	2.4634

(a) Assuming (x_1, x_2) to have a multivariate normal distribution with same dispersion matrix in the two populations, test whether the two types differ with respect to x_1, x_2.

(b) Assuming x_1, x_2 to be independent, test the same hypothesis.

(c) Test whether x_1, x_2 are uncorrelated in the set-up of (a).

(d) Assuming the sample dispersion matrix as equal to the true common dispersion matrix, test the hypothesis formulated in (a). (8+6+6+4)=24

5.(a) A sample of size 100 was drawn at random from a normal population with mean = variance = θ. The sample mean and variance were 9 and 10 respectively. Find the maximum likelihood estimate of θ and estimate its variance. (12)

(b) $x_1, x_2, \ldots$ are a sequence of independent and identically distributed random variables with density $f(x|\theta) = \theta e^{-x\theta}$, $x \geq 0$. You are to test $H_0(\theta = 1)$ against $H_1(\theta = 2)$ with $\alpha = \beta = .01$. Find the boundaries for $\sum_1^n x_i$, if you use an SPRT for this purpose. Calculate the ASN of your test under $H_0(\theta = 1)$ and under $\theta = 1.5$. Compare with the best non-sequential test which guarantees $\alpha = \beta = .01$. (Excess over boundaries may be neglected.) (10+14)=24

Neatness (4)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper I : Official Statistics and Descriptive Statistics (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks.

GROUP A

(Answer any three questions from this group)

1. You are required to compare the rates of growth of industrial and agricultural production in India, for the post-independence period by certain major categories of output. Describe the type of data you would use for the purpose, mentioning the publications you would consult. Do the publications mentioned by you provide data separately for the States also? (16)

2. Mention the main sources of statistics for India in any two of the following fields:

   i) Foreign trade iii) National income

   ii) Education iv) Population and its rate of growth

   Sketch the broad contents of the relevant publications and indicate the uses of the information. Comment on the major limitations of the available data. (6+10)=16

3. Describe any one of the Indian Sample Surveys of current interest, indicating the purpose of the survey, nature of the data collected, the general sample design and sample size, some uses of the data and reliability of the information collected. (16)

4.(a) Describe briefly the main functions, in the field of statistics, of any one of the following agencies :-

   i) Central Statistical Organisation;

   ii) Reserve Bank of India

   iii) Indian Statistical Institute

   (b) What do the following abbreviations stand for :

   ILO, FAO, UNESCO, WHO, UNICEF, GATT, ECAFE, and OECD ?

   Indicate briefly the functions of any one of the above agencies. (8+4+4)=16

5.(a) Mention the agencies or offices you will approach for obtaining information on any three of the following :

   i) index number of mineral production

   ii) hydro-electricity generated in public utilities

   iii) value of import licenses issued

   iv) number of radio receivers in use

   v) industry-wise despatches of coal

Please turn over

5.(b) Describe briefly the statistical reporting system on which the primary data used for obtaining any two of the following estimates are based :

i) production of wheat in Uttar Pradesh

ii) production of finished steel in India

iii) amount of income-tax collected during a year

iv) contribution to national income by distribution trades

(6+10)=16

Neatness (2)

GROUP B

(Answer any three questions from this group)

6.(a) Distinguish between primary data and secondary data. What precautions are necessary before making use of the secondary data?

(b) It is proposed to investigate the extent of participation in religious and social activities of a community residing in a certain locality in a large city. Explain how you will define the population and the sampling unit. State clearly the type of sampling frame you would use to design a sample for the purpose.

(c) Describe , in brief, the various stages involved in processing statistical data and the steps you would take to control the errors in processing work. (5+5+6)=16

7.(a) Show that the point of intersection of the ogives of 'less than' and 'greater than' type corresponds to the median value.

(b) Explain the circumstances in which the various measures of central tendency are used giving one illustration for each measure.

(c) For a distribution of 280 heights, mean = 54" and standard deviation = 3". On checking, it was discovered that two heights which should correctly read as 62" and 82", had been wrongly recorded as 64" and 80" respectively. Calculate the correct values of the mean and standard deviation. (3+8+5)=16

8.(a) Derive a hypergeometric distribution and find its mean and variance.

(b) For a binomial distribution, $(q+p)^n$, where $q = 1-p$, show that

$$k_{r+1} = pq \frac{dk_r}{dp}, \quad r \geq 1,$$

where k_r denotes the rth cumulant. (8+8)=16

9.(a) If $u = ax + by$, $v = bx - ay$ and u and v are uncorrelated, prove that

$$\sigma_u \sigma_v = (a^2 + b^2)\, \sigma_x \sigma_y \sqrt{1 - \rho^2}$$

where ρ is the correlation coefficient between x and y.

(b) For the variables x and y, the equations of the 2 regression lines are

$4x - 5y + 33 = 0$ and $20x - 9y - 107 = 0$

Identify the regression line of y on x and that of x on y. What is the estimate of y when $x = 10$?

If this estimate is denoted by y_0, find the estimate of x when $y = y_0$. Is this estimate equal to 10? Justify your statement. (6+1+2+7)=16

10.(a) Prove that

$$R^2_{1(23)} \geq \rho^2_{12}$$

(b) If all the correlation coefficients of zero order in a set of k variables are equal to ρ, show that every partial correlation of s^{th} order is $\frac{\rho}{1+s\rho}$ (7+9)=16

Neatness (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper II : Probability Theory and Statistical Methods (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks.

GROUP A

(Answer any four questions from this group)

1.(a) Show that for any n events $A_1, \ldots, A_n$

$$P(A_1 \cup \ldots \cup A_n) \leq P(A_1) + \ldots + P(A_n)$$

(b) Suppose $P(A) = 0.2$, $P(B) = 0.3$ and $P(C) = 0.4$

Assuming A, B and C to be independent, find $P(A \cup B \cup C)$.

(c) A balanced die is thrown twice. Let X and Y denote the number of points observed on the first and the second throw respectively. Given that $X + Y$ is even, find the probability that X is greater than Y. (4+4+4)=12

2.(a) Define probability generating function.

(b) Let $m > 0$ be a constant and $p(x,m) = C\, m^x/(x!)$, $x = 0, 1, 2, \ldots$.

Determine the value of the constant C which will make the above function $p(x, m)$ a probability distribution.

(c) Let X_1, X_2 be independent random variables having the distribution $p(x, m)$ given in (b). Find the distribution of $(X_1 + X_2)$.

(2+4+6)=12

3. For the joint probability density

$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < x < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

find

i) the marginal cumulative distribution functions $F(x)$ and $G(y)$;

ii) the marginal probability densities $f(x)$ and $g(y)$;

iii) the conditional probability densities $f(x \mid y)$ and $g(y \mid x)$.

(4+4+4)=12

4. Let X, Y be random variables having finite positive variances.

(a) Define the correlation coefficient ρ between X and Y

(b) With ρ as in (a), prove that $-1 \leq \rho \leq 1$

(c) Show that $\rho = 0$ if X and Y are independent.

(d) Give an example where $\rho = 0$ but X and Y are not independent.

(2+4+3+3)=12

Please turn over

5.(a) State carefully any version of the central limit theorem.

(b) Using the normal approximation to the binomial, estimate the number of times that a coin should be tossed to make it at least 95% sure that the observed proportion of heads is within .01 of the true (but unknown) probability p of heads.

[You may use the following facts : (i) $p(1-p) \leq 1/4$; (ii) the probability is 0.95 that a standard normal random variable lies between -1.96 and +1.96] (4+8)=12

Neatness (2)

GROUP B

(Answer any four questions from this group)

6. Let $x_1, x_2, \ldots, x_n$ be independent observations on a random variable X which assumes the values 0 and 1 with probabilities $(1 - \pi)$ and π respectively.

(a) Show that

$$r = \sum_{1}^{n} x_i$$

is a sufficient statistic for the parameter π.

(b) Derive the maximum likelihood estimator of π. (6+6)=12

7.(a) What is meant by the statement: "the estimator $\delta(x)$ of the parameter θ is admissible with respect to the squared error loss"?

(b) Show that $a\bar{x}$ is inadmissible for E(X), with $|a| > 1$ and with squared-error loss. (6+6)=12

8.(a) State and prove the Neyman-Pearson lemma.

(b) Let $x_1, x_2, \ldots, x_n$ be a random sample of size n drawn from a population with density function

$$\theta \exp(-\theta x); \quad x > 0, \quad \theta > 0$$

Derive the most powerful test for testing $H_0(\theta = \theta_0)$ against $H_1(\theta = \theta_1 > \theta_0)$ and show that the test is actually uniformly most powerful against one-sided alternatives. (6+1+2)=12

9.(a) What is the difference between a regression problem and an analysis of variance problem ?

(b) How would you test the hypothesis that two regression lines have the same slope? (4+8)=12

10.(a) Discuss the approximations that may be used in the case of large samples.

(b) What is meant by standard error of a statistic? Account for its importance in large-sample estimation and large-sample tests of significance. (4+1+1)=12

Please turn over

(a) Let a population be classified according to two attributes - A and B, into K and l classes respectively. Let π_{ij} be the proportion of members of the population belonging simultaneously to the ith class of A and the jth class of B. How would you test whether A and B are independent, when π_{ij}'s are unknown?

(b) One evening, three persons suspected to be driving under the influence of liquor were stopped and blood samples taken from each were sent to the laboratory. Five determinations of the percentage of alcohol in the blood were made on each sample. State how you will apply the technique of analysis of variance to test whether all three drivers were equally intoxicated.

(6+6)=12

Neatness (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper III : Sample Surveys and Design & Analysis of Experiments (Theoretical)

Time : 4 hours Full marks : 100

Figures in the margin indicate full marks

GROUP A

(Answer any three questions from this group)

1. Discuss briefly the following statements :

(a) Even when complete inspection is possible, sampling may have economic advantages.

(b) When only a very small sample can be afforded, a judgement sample is often preferable to a random sample.

(c) It is a remarkable fact that the measure of precision of the estimate can itself be estimated from the random sample, without having to know the true value of the estimate.

(d) For populations with similar standard deviations, the larger the populations size N, the smaller will be the sampling fraction $\frac{n}{N}$, required to attain the same precision in the estimate. (4+4+4+4)=16

2.(a) What is stratified sampling? Which are the situations under which stratified sampling is advantageous?

(b) Find the optimum allocation to k strata using a cost function of the type

$$C = C_0 + \sum_{i=1}^{k} n_i c_i ,$$

when simple random sampling without replacement is adopted within each stratum.

What are the practical difficulties in using the optimum allocation?

(c) When simple random sampling with replacement is used within strata, compare the efficiencies of proportional and optimum allocations for a fixed sample size.

3.(a) How would you use the information on an auxiliary variable x, correlated to the study variable y, to improve the simple unbiased estimate of the total of y in simple random sampling without replacement?

(b) Derive the approximate expressions for the bias and mean square error of a ratio estimator, stating clearly the assumptions involved. (4+12)=16

4.(a) What are the sources of non-sampling errors? How do these errors behave with increase in sample size?

(b) Describe briefly the techniques available for assessing non-sampling errors.

(c) What steps would you recommend to control non-sampling errors? (5+5+6)=16

Please turn over

5. Give, for each of the following, an example of an actual problem and the circumstances where you would most certainly not recommend the use of the sampling technique indicated below. State clearly your reasons.

a) systematic sampling

b) cluster sampling

c) sampling the same units on successive occasions

d) cumulative method for sampling with unequal probabilities (4+4+4+4)=16

Neatness (2)

GROUP B

(Answer any three questions from this group)

6.(a) Discuss briefly the role of randomization, replication and local control in designing statistical experiments.

(b) Describe clearly a randomized block design. (10+6)=16

7. What is meant by 'missing-plot technique' ?

Suppose the yield of the i-th treatment in the j-th row and k-th column in an n X n Latin square is missing. Describe clearly how you would analyse the data. (4+6+6)=16

8. Explain what is meant by a simple lattice design. State (without proof) the relations between its parameters. What is the essential difference between this design and the balanced incomplete block design? Obtain expressions for the estimates of treatment effects and also of the sum of squares due to treatments. Derive the efficiency of the design. (2+2+1+3+3+5)=16

9.(a) What are the situations where adjustment for concomitant variation is called for?

(b) Derive a suitable computational technique and indicate what tests should be carried out in the problem of testing for differences in growth caused by different feeding treatments, when the average initial weights of groups of individuals chosen for different treatments are possibly different. (4+12)=16

10. It is desired to estimate by using an incomplete block design elementary treatment contrast with equal accuracy, while eliminating block effects and positional effects inside blocks. Indicate which design you will use for this purpose. Write down the analysis of variance table for this design. Also obtain expressions for the estimates of treatment effects. (4+5+7)=16

Neatness (2)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1960

Paper IV : Applied Statistics (Theoretical)

Time : 4 hours Full marks : 100

(a) Candidates will be required to answer questions from those two groups of subjects only, for which they have registered their option.

(b) Separate answer books are to be used for each of the two groups attempted.

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is not permitted.

GROUP A : ECONOMIC STATISTICS

(Answer any three questions from this group)

1. Give an outline of the moving average method of determining constant seasonal indices from a time series, with computational layouts and explanations. Why is the method preferred to the method of ratio-to-trend ?

(12+4) = 16

2. Starting from first principles, derive the equation of the concentration curve of the lognormal distribution, and bring out the symmetry property of this concentration curve.

(12+4) = 16

3. Describe the open input-output table for an economy and state the basic assumptions underlying its major applications. In what way is input-output analysis different from the usual production function analysis ?

(6+5+5) = 16

4. (a) What is a demand function ? Explain fully how family budget data enable us to study the individual consumer's demand as a function of his income - all other factors affecting demand remaining approximately constant.

(b) Explain clearly the concept of elasticity of demand with respect to various factors. Indicate the range of values assumed by each elasticity for various classes of commodities.

(3+7+6) = 16

5. (a) Explain clearly the distinction between national income at factor cost and national income at market prices.

(b) Write a critical note on the weaknesses of the official national income estimates currently available for India.

(6+10) = 16

6. Write short notes on any three of the following :

i) The time reversal test of index number formulae

ii) Various types of errors in index number calculations

iii) Demand projections based on Engel curves

iv) The problem of technological change in the estimation of production functions from time series data

v) Growth of India's national income during the Five-Year Plans targets versus achievements.

(16)

Neatness (2)

GROUP D : STATISTICAL QUALITY CONTROL

(Answer any three questions from this group)

1. (a) Briefly describe the role of the Poisson Law in quality control and list down four industrial examples where it holds good.

(b) Explain how a control chart helps in controlling a measurable characteristic during a production process. Describe how to construct control charts for,

i) number defectives
ii) defects per unit

when sample size varies. (6+10) = 16

2. Obtain an unknown sigma acceptance sampling plan by variables for the case of one-sided specification (upper) limit U, such that under the plan, lots with $100p_1$ and $100p_2$ per cent defectives ($p_2 > p_1$) would be accepted with probabilities $(1-\alpha)$ and β respectively. Write out the expression for the O.C. function of the plan. (12+4) = 16

3. Distinguish between the natural tolerance limits and the control limits.

In the following two cases, derive the necessary expressions to construct tolerance limits such that it can be asserted with 100β per cent confidence that they will include at least 100α per cent of the population :

i) a normal parent population with known mean

ii) a parent population with unknown form,

where α and β are preassigned numbers between 0 and 1. (2+7+7) = 16

4. (a) Write down the salient features of Mil-Std 105D acceptance sampling plan by attributes and indicate its distinction from Mil-Std 105B.

(b) Certain items were submitted in batches for inspection under Mil-Std 105D. In the early part of the period of inspection, reduced inspection remained in vogue and during the latter part normal inspection was invoked, according to the rule of the Mil-Std 105D. During the period of inspection no batch was rejected and the process remained steady.

If the plans used above were as follows :

type of inspection	level	AQL	sample size	acceptance number	rejection number
Reduced	II	p	n_1	c_1	r_1
Normal	II	p	n_2	c_2	r_2

obtain expected inspection cost per batch during the above period, assuming cost of inspection per item as Rs. C.

(10+6) = 16

Neatness (2)

GROUP C : STATISTICAL METHODS IN GENETICS

(Answer any three questions from this group)

1. For detecting linkage between two factors segregating into
(a) 1:1 and 1:1 (double backcross) ratio,
and (b) 3:1 and 3:1 (F_2) ratio,
obtain the suitable forms of χ^2 test and estimate the linkage values ($\hat{p}$) in each case by the maximum likelihood method. Also find out the values of the standard error of these estimates. (15)

2. Under assumption of absence of differential fertilization or viability or linkage, obtain the following estimates for the (n+1) generation when F_1 plants are selfed for 'n' successive generations :
(a) Variance of F_{n+1} progenies,
(b) Variance of means of F_{n+1} progenies,
(c) Mean variance of F_{n+1} progenies, and
(d) Covariance of F_{n+1} progeny means and F_n parents. (16)

3. Prove that under mass selection in a random mating population, the gain expected per generation is equal to the product of the degree of heritability and the selection differential.

Describe a method for constructing selection indexes for multiple objectives. (16)

4. By using the maximum likelihood method, obtain estimates of the blood-group gene-frequencies of mothers and children from 'mother-child' combinations of 'ABO' blood-groups; also obtain the variances of these estimates.

What are the limitations of this method ? (16)

5. Write short notes on any three of the following :
i) Coefficient of heritability under random mating and inbreeding;
ii) Proband method of ascertainment in human genetics;
iii) Principles of Discriminant Functions in plant selections;
iv) Methods of artificial selection in random mating population and effects on gene-frequencies. (16)

Neatness (2)

GROUP D : VITAL STATISTICS AND DEMOGRAPHY

(Answer any three questions from this group)

1.(a) Give a critical account of the 1961 Census questions.

(b) What improvements would you suggest to be incorporated in the questionnaire for the 1971 Census ? (16)

2.(a) Discuss the different methods of projection of the population of a community.

(b) Give a suitable formula for finding the population of India in 1969. (16)

3.(a) Discuss the importance of finding the infant mortality rate and its sub-divisions by duration of life of baby. State the limitations of this rate.

(b) Write a note on the standardized death rate. (16)

4. i) What are the assumptions made while applying a life-table (derived from a given population) for practical purposes ?

ii) Describe the structure of a complete life table.

iii) Explain how a complete life table may be constructed from vital statistics and census returns.

(3+5+8) = 16

5. Write notes on any three of the following :

(a) Vital registration in India

(b) Logistic curve

(c) Measures of morbidity

(d) Age of female at marriage as a factor affecting fertility

(e) Working population (16)

Neatness (2)

GROUP B : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer question 1 and any two questions from this group)

1. What do you mean by the reliability of a test ? What are the methods of estimating reliability ? Can a test be such that it has high validity but low reliability ? If so what is the nature of such a test ?

(4+8+2+4) = 18

2. Describe different types of validity of a test. Derive the formula which shows the relation among the reliabilities of the test and the criterion, the original validity coefficient and the maximum validity the test can attain, when both the criterion and the test are perfectly reliable.

(8+8) = 16

3.(a) Describe how you could obtain the confidence interval (at $\alpha \leq$ level) of the true score corresponding to a given observed score.

(b) What do you mean by the point-biserial correlation coefficient of an item ? Derive the formula used for estimating this coefficient.

(6+10) = 16

4.(a) Illustrate geometrically the following :

i) inter-correlations among the tests
ii) common factor loadings
iii) the concept of "communality".

(Consider the case when there are three tests and two orthogonal common factors)

(b) What do you mean by (i) Specific Factor and (ii) Group Factor ?

(12+4)=16

5. Write short notes on any four of the following :

(a) Scoring formulae

(b) Parallel Tests

(c) Spearman-Brown Formula

(d) Standard Error of Measurement

(e) Rank Order Correlation

(4+4+4+4) = 16

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper V : Methods of Numerical Computation; Descriptive Statistics and Official Statistics (Practical)

Time : 5 hours Full marks : 100

(i) Figures in the margin indicate full marks.

(ii) Use of calculating machines is permitted.

GROUP A

(Answer any two questions from this group)

1.(a) Given

$f(0) = 1.056$

$f(1) + f(2) = 10.973$

$f(3) + f(4) + f(5) = 65.686$

find the value of $f(6)$.

(b) The population of India, as recorded during the decennial censuses, is given below :

Census year	Population (in lakhs)
1901	2,384
1911	2,522
1921	2,514
1931	2,791

Find, by sub-division of intervals, the population corresponding to each of the years between 1901 and 1911. (6+6)=12

2.(a) For what value of x is the following tabulated function a minimum ?

x	f(x)
0.2	0.9182
0.3	0.8975
0.4	0.8873
0.5	0.8862
0.6	0.8935
0.7	0.9086

(b) Find, to four decimal places, the real root of the equation $x^3 + x - 1 = 0$ by any method you know. (6+6)=12

3.(a) The corresponding values of the function U_x are given for the following values of x. Use Lagrange's formula to obtain the value of U_{19}.

x	U_x
11	14,646
17	83,526
21	191,486
23	279,846
31	923,526

Please turn over

3.(b) Given the following table of values of $U_{x:y}$, estimate the value of $U_{23:17}$:

x \ y	15	20	25
20	5.917	4.418	3.547
25	6.046	4.530	
30	6.144		

(6+6)=12

GROUP B

(Answer all four questions from this group)

4. In the following frequency table, x denotes the deviation of mid-point of a class-interval, from an arbitrary origin expressed in terms of the constant width 'h' of class-interval as unit, and f the frequency of the class-interval values.

x	-3	-2	-1	0	1	2	3	4
f	3	15	43	57	50	36	25	9

Find 'h' and the boundaries of all the class-intervals, if the mean and standard deviation of the distribution are respectively 40.604 and 7.92. (10)

5. The mean of a normally distributed variable x is 2.08, and 9.68% of the values of x are negative. Find the probabilities that a value of x taken at random will be –

(a) greater than 5.6,

(b) less than 1.8,

and (c) greater than 1.8 but less than 5.6. (10)

6. An experiment has been conducted to verify the law of falling bodies under gravity expressed by the equation $s = \frac{1}{2}gt^2$, where s is the distance fallen at time t and g is the gravitational constant.

The following results have been obtained :

t (seconds)	1	2	3	4	5
s (feet)	15	70	140	250	380

Using the data quoted above, estimate the value of the constant 'g'. (10)

Please turn over

7. EITHER

(a) The table below gives the annual crop cutting data of a certain place during the period 1919-20 to 1945-46.

year	yield per acre (lbs.)	condition factor
(1)	(2)	(3)
1919-20	100.5	114.0
20-21	80.0	45.0
21-22	109.2	105.0
22-23	91.8	82.5
23-24	80.9	82.5
24-25	77.3	82.5
25-26	71.4	67.5
26-27	78.7	67.5
27-28	103.0	90.0
28-29	111.2	93.8
29-30	101.3	87.0
30-31	101.1	87.0
31-32	39.8	41.2
32-33	82.0	69.0
33-34	71.0	65.3
34-35	58.8	55.5
35-36	60.6	61.5
36-37	86.1	78.0
37-38	69.0	69.0
38-39	60.9	57.0
39-40	86.5	85.5
40-41	99.7	97.5
41-42	111.8	101.2
42-43	58.2	63.7
43-44	89.3	75.7
44-45	56.4	60.7
45-46	67.9	65.6

Calculate trend values by using a five year moving average.

By arranging to place the moving averages in such a manner as would make the last average correspond to the 1945-46 figure, show that the condition factor series serves as a very reliable guide for estimating the yield-per-acre values.

Estimate also the yield values for the last two years. (18)

OR

(b) From the results of a crop-cutting experiment the following values were obtained for 15 cinchona plants :

plant no.	yield of bark (ounces)	height of plant (inch)	girth of plant (inch)	number of stems in the plant
(1)	(2)	(3)	(4)	(5)
1	105	133	8	1
2	34	95	4	1
3	22	62	2	1
4	56	118	4	1
5	60	109	5	1
6	50	74	3	1
7	183	167	5	2
8	22	98	3	1
9	74	78	4	2
10	56	107	3	3
11	8	47	1	1
12	53	129	4	1
13	45	115	4	1
14	19	92	3	1
15	39	120	5	1

How far do the last three factors considered, influence the yield of cinchona bark in the given experiment ? (18)

Please turn over

GROUP C

(Answer both the questions from this group)

8. From the official publications placed at your disposal, attempt any two of the following questions :

(a) Write down the contribution to national income by agriculture, animal husbandry and ancillary activities at 1948-49 prices for 1950-51 and 1960-61. Write down also the index number of agricultural production for the two years and compare the two rates of growth.

(b) Write down the general index of industrial production for five latest years, side by side with the wholesale price index for manufactured articles. Which have increased more - the production or prices ?

(c) Express the number of school students at nursery, primary and secondary stages in 1960-61 as percentages of 1961 population. $(7\frac{1}{2}+7\frac{1}{2})=15$

9. From the official publications furnished, tabulate neatly the information required for any three of the following cases :

(a) The All-India Census male and female population of 1961, split up under the various economic classifications available [figures under classification details to be quoted in millions only correct to one place of decimal].

(b) Number of scholars (men and women separately) studying under the various types of colleges for professional and special education, for any 2 recent available years.

(c) Number of wagons loaded commodity-wise with Revenue traffice only, on the Broad Gauge of Indian Government Railways, during any two recent available months.

(d) Index of consumer prices for any two recent available months

for i) Industrial workers in Ahmedabad, Bombay, Kanpur and Calcutta

and for ii) Agricultural labour in Andhra Pradesh, Bihar, Maharashtra and Uttar Pradesh.

(e) i) Quantity of electricity generated and sold (the quantity sold is to be reported under the various heads available).

ii) Quantity of Cotton piece goods (cloth) produced, by the categories available

for any 2 recent months in both cases.

[Please note that the appropriate units and base years are to be shown whenever necessary. Also indicate without fail the page references of the publications consulted.] (9)

Neatness (Group A + Group B + Group C) (4)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper VI : Statistical Methods; Design & Analysis of Experiments and Sample Surveys (Practical)

Time : 5 hours | Full marks : 100

(a) Figures in the margin indicate full marks.

(b) Use of calculating machines is permitted.

GROUP A

(Answer any two questions from this group)

1.(a) Controlled studies performed on the response of Sharks to human perspiration, gave the following data :

type of shark	response: strong repulsion	mild repulsion	neutral	total
(1)	(2)	(3)	(4)	(5)
Black tip	13	20	6	48
Others	8	20	15	43
Total	21	40	21	91

i) Test for no association between the nature of response and the type of shark.

ii) Test the hypothesis that $\leq 25\%$ of black tip sharks are strongly repelled by human perspiration against the alternative that more than 25% are so repelled.

iii) Find a 90% confidence interval for the true proportion of black tip sharks which remain neutral to human perspiration.

(b) Of 101 earthquakes that occurred in a country over twenty years 27 occurred in spring, 27 in summer, 18 in autumn and 29 in winter. Test whether occurrence of earthquakes in the country may be considered to vary from season to season. (6+4+4+6) = 20

2.(a) Given the following observations on a variate y at different levels of an auxiliary variable x, fit the regression line $Y = \alpha + \beta x$ by the method of least squares.

x :	4.5	5.0	5.5	6.0	6.5	7.0	8.0	8.5	9.0
y :	14.0	13.4	13.2	13.1	14.9	15.8	15.0	16.0	16.7

Test whether the population β can be taken to be 0.6. Also set up a suitable interval which would include any observation on the variate y corresponding to $x = 9.5$, with 90% confidence.

(b) The following correlation matrix was obtained from a study of four factors X_1, X_2, X_3, X_4 affecting the volume Y of bread on the basis of a sample of size 22.

X_1	X_2	X_3	X_4	Y
1	.3090	-.4528	.5085	.5389
	1	-.2313	.9380	.5377
		1	-.4256	-.4784
			1	.6134
				1

Test whether in predicting Y from the factors with the help of multiple linear regression, it is wholly unnecessary to include the predictor X_4 over and above X_1, X_2, X_3. (4+1+1+8)=20

Please turn over

3.(a) The following two-way classified data were obtained from a study of the effects of well-defined levels of exercise, on the blood cholesterol content of 15 individuals in two age groups. Make a complete analysis of the data.

age group	level of exercise		
	low	medium	high
(1)	(2)	(3)	(4)
20 - 29	190	155	155
	210	165	150
	170	190	155
30 - 39	200	170	175
	210	190	155

(b) Two research workers A and B were asked to investigate whether a physical measurement for adult members of a certain tribe was on an average greater for women than for men. Investigator A proceeded by taking independent random samples of 30 men and 30 women, whereas B selected randomly 30 families and for each family took measurements both on the husband and the wife. The results are summarised below (x stands for observation on a man and y for that on a woman) :

Investigator A

$\Sigma x = 25.31$ $\Sigma y = 25.96$

$\Sigma x^2 = 22.21$ $\Sigma y^2 = 22.94$

Investigator B

$\Sigma x = 25.44$ $\Sigma y = 26.16$

$\Sigma x^2 = 22.12$ $\Sigma y^2 = 23.06$

$\Sigma xy = 22.46$

With the help of appropriate tests, see what conclusions follow from the sets of data obtained by the two investigators. Combine the evidences of the two tests and comment on the postulated hypothesis. (10+5+5)=20

GROUP B

(Answer any two questions from this group)

4. The following design was used to test 9 rations fed to rats. The gains in weight of the rats after the feeding experiment were as follows : (The ration numbers are in brackets)

Replication 1

Block 1	(1) 20	(4) 15	(7) 11
" 2	(3) 08	(6) 18	(9) 23
" 3	(2) 18	(5) 16	(8) 02

Replication 3

Block 7	(1) 13	(9) 19	(5) 14
" 8	(8) 14	(4) 34	(3) 02
" 9	(6) 14	(2) 20	(7) 14

Replication 2

Block 4	(7) 08	(8) 12	(9) 16
" 5	(1) 20	(2) 02	(3) 02
" 6	(4) 20	(5) 06	(6) 02

Replication 4

Block 10	(5) 10	(7) 23	(3) 06
" 11	(1) 22	(6) 12	(8) 02
" 12	(9) 27	(2) 07	(4) 20

Analyse the data and compute the efficiency of this design relative to randomised block design. (12+3)=15

Please turn over

5. The table below gives the plan and yields of turnips in cwt. per acre, of two replications of a 3^3 experiment, to test nitrate, phosphate and potash dressings in 6 blocks of 9 plots each. The levels of the factors were as follows :

n_0, n_1, n_2 are 0, 30 and 60 lbs. of N per acre;

P_0, P_1, P_2 are 0, 40 and 80 lbs. of P_2O_5 per acre;

k_0, k_1, k_2 are 0, 50 and 100 lbs. of K_2O per acre respectively.

Identify the confounded effect or effects and carry out a complete statistical analysis. The treatment combination $n_i p_j k_n$ is denoted by ijn were i, j, n = 0, 1, 2. For example 112 denotes the treatment combination of nitrate, phosphate and potash at the levels 1, 1 and 2 respectively.

Replication I

Block I		Block II		Block III	
Treatment	Yield	Treatment	Yield	Treatment	Yield
(1)	(2)	(3)	(4)	(5)	(6)
110	11	200	42	210	46
121	11	120	20	001	48
212	13	021	24	012	58
000	12	010	38	221	53
220	11	112	39	100	54
102	30	222	61	122	37
011	41	002	40	020	41
201	21	101	46	202	25
022	21	211	44	111	32

Replication II

Block I		Block II		Block III	
Treatment	Yield	Treatment	Yield	Treatment	Yield
(1)	(2)	(3)	(4)	(5)	(6)
001	39	200	32	212	26
202	34	002	37	220	16
122	31	222	37	011	06
100	34	010	33	110	13
012	40	021	27	201	12
020	31	101	21	022	12
111	26	211	12	121	16
221	35	120	30	102	19
210	38	112	22	000	14

(3+12)=15

Please turn over

6. A latin square experiment was carried out to test for possible differences between breaking strengths of glass (in lbs. per square cm.) made from six furnaces using six different moulds. Six runs of each furnace were needed to complete the 36 possible combinations of furnaces and moulds. One particular observation was spoilt through the cooling apparatus going out of order and has been shown as 'blank' in the table of breaking strengths below :

(The letters refer to the six moulds)

	FURNACES					
	1	2	3	4	5	6
1	B 103	E 111	D 105	C 118	A 101	F 130
2	D 113	B 104	E 115	A 106	F 107	C 129
3	A 125	D 108	B 102	F 104	C 103	E 119
4	F 109	A 123	C 106	D –	E 103	B 111
5	E 106	C 107	F 103	B 114	D 113	A 117
6	C 102	F 110	A 108	E 104	B 115	D 131

Analyse the data, giving means and standard errors. (10+2+3)=15

GROUP C

(Answer any one question from this group)

7. The following table relates to data from a complete census of 2010 farms carried out in a County in America. The farms were arranged in seven size-groups.

stratum no.	farm size (acres)	number of farms	total area under corn (acres)	standard deviation per farm (acres)
(1)	(2)	(3)	(4)	(5)
1	0 – 40	394	2127	8.3
2	41 – 80	461	7402	13.3
3	81 – 120	391	9515	15.1
4	121 – 160	334	11524	19.8
5	161 – 200	169	7110	24.5
6	201 – 240	113	5651	26.0
7	241 and above	148	9438	35.2

It is required to plan for a stratified selection scheme, in order to estimate the mean acreage per farm under corn in a certain year. Find out the best possible distribution of sample farms in the various strata when the total cost of the survey is fixed at 2,700 dollars, and when the cost of investigation and

Please turn over

7. (contd.) laboratory analysis per farm varies from stratum to stratum as follows :

stratum number	cost per farm (dollars)
1	25.5
2	27.6
3	28.0
4	29.3
5	29.8
6	30.5
7	31.6

Calculate also the variance of the final estimate.

Suppose a new system of strata is formed combining the old strata as follows :

old stratum number	new stratum number
1 and 2	1
3 and 4	2
5 and 6	3
7	4

Calculate the variance and cost under the optimum allocation for the new scheme of strata, when a total sample of 100 farms is taken. (15+12+3)=30

8.(a) The population values of a variable are shown below, in such a way as to facilitate the drawing of a systematic sample with N = 40, K = 10 and n = 4.

strata	systematic sample numbers									
	1	2	3	4	5	6	7	8	9	10
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
1	0	1	1	2	5	4	7	7	8	6
2	6	8	9	10	13	12	15	16	16	17
3	18	19	20	20	24	23	25	28	29	27
4	26	30	31	31	33	32	35	37	38	38

Calculate the variance of the systematic sample mean and compare it with the variance in the cases of unstratified

(i) simple random sampling

and (ii) stratified random sampling.

If the order of observations in the 3rd and the 4th strata are reversed what are the changes in the above results.

Please turn over

8.(b) A sample survey was conducted in 1946 in a district for estimation of the total number of cattle-heads. The following data give the numbers of cattle-heads enumerated, together with those given in the 1945 Livestock Census, in 15 villages selected with equal probability and without replacement.

The total number of villages in the district is 135 and the total cattle population according to the 1945 census is 14329.

number of cattle heads (1946) y	number of cattle heads (1945) x
(1)	(2)
634	623
696	690
530	534
315	293
078	069
640	842
692	475
292	371
210	161
533	298
2110	2045
592	1069
707	706
375	330
212	218

Compare the ratio method of estimation with the conventional unbiased estimation for estimating the total number of cattle heads in 1946.

You are furnished with the following information on uncorrected sums of squares and products :-

$$\sum_{i=1}^{15} x_i^2 = 8406598, \qquad \sum_{i=1}^{15} y_i^2 = 8112176$$

$$\text{and} \quad \sum_{i=1}^{15} x_i y_i = 8060397 \qquad (20+10)=30$$

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper VII: Applied Statistics (Practical)

(Two groups only are to be attempted)

Time 3 hours Full marks: 100 (covering two groups of 50 marks each)

(a) Candidates will be required to answer questions from those two groups of subjects only, for which they have registered their options.

(b) Separate answer books are to be used for each of these two groups attempted.

(c) Figures in the margin indicate full marks.

(d) Use of calculating machines is permitted.

GROUP A : ECONOMIC STATISTICS

(Answer all three questions from this group)

1. The following data relate to six main crops of a certain country. By calculating price and quantity indices for the year 1939, taking 1929 as base, show whether Paasche type index numbers meet the factor reversal test.

crop	unit	average farm prices ($)		annual production (million units)	
		1929	1939	1929	1939
corn	bu.	0.774	0.485	2516	2581
cotton	lbs.	0.164	0.091	7415	5910
hay	ton (sh.)	12.19	7.61	76.02	76.38
wheat	bu.	1.035	0.777	824.2	741.2
oats	bu.	0.426	0.334	1113	958
potatoes	bu.	1.288	0.700	333.4	342.4

(17)

2. The following shows the general index of industrial production in India (1950=100). Fit an exponential trend to the series.

Year	1956	1957	1958	1959	1960	1961	1962	1963	1964
Index	100.0	104.2	107.7	116.9	130.1	138.3	150.5	162.5	174.7

(18)

3. EITHER

(a) Use the following data based on the 14th round of the National Sample Survey (NSS), to estimate the parameters of the semi-log form of Engel curve for cereals by weighted least squares method. Hence obtain the Engel elasticity of cereals at the mean level of total consumer expenditure per capita.

Table : Estimates for Rural India : NSS 14th Round

monthly per capita total consumer expenditure (Rs.)	percentage of population	average monthly per capita expenditure (Rs.)	
		cereals	all items
0-8	9.03	3.68	6.18
8-11	14.73	5.57	9.49
11-13	11.33	6.86	12.01
13-15	9.87	7.63	13.08
15-18	13.05	8.51	16.44
18-21	9.48	9.21	19.53
21-24	8.49	10.41	22.46
24-28	6.96	10.37	25.66
28-34	6.17	11.49	30.34
34-43	5.34	12.82	38.26
43-55	2.75	14.32	48.23
55-	2.80	16.49	89.45

OR

3.(b) i) Test graphically whether the size distribution of per capita total consumer expenditure presented in Q3 (Either) is approximately lognormal or not.

ii) Utilize the same data to compute the concentration coefficient of per capita total consumer expenditure, without making any assumption regarding the form of the underlying distribution. (12+12)=24

GROUP B : STATISTICAL QUALITY CONTROL

(Answer any two questions from this group)

1. The following table gives the mean values and ranges, in 30 samples (of size 4 each), of percentage carbon content in commercial metal sheets, as obtained from test records.

The data were collected for a continuous period of 30 days assuming that the percentage carbon content within a day is unlikely to show high variation from sheet to sheet.

Sample	Average	Range	Sample	Average	Range
1	11.10	0.6	16	11.45	1.3
2	11.70	1.2	17	11.55	1.6
3	11.35	1.0	18	9.98	0.4
4	11.25	1.0	19	10.78	1.2
5	11.40	2.0	20	11.23	0.7
6	11.00	0.6	21	10.93	1.7
7	11.20	1.0	22	11.50	2.7
8	11.35	1.2	23	10.78	0.7
9	11.50	2.0	24	10.95	1.1
10	10.88	1.1	25	11.48	2.9
11	10.85	1.0	26	10.80	0.4
12	11.53	1.2	27	12.20	2.0
13	11.15	0.8	28	11.88	1.5
14	11.28	1.0	29	11.23	0.8
15	11.00	0.8	30	11.30	0.6

(a) Test by means of control chart whether the process is under statistical control.

(b) A minimum of 8% carbon in any sheet is the market specification. Excess of 0.2 per cent on an average over the market specification results in a loss of Rs. 9,000 per annum to the factory. Estimate how much saving can be effected by maintaining statistical control at a proper level satisfying market specification.

(15+10)=25

2. Under an acceptance-rejection inspection scheme items were selected one by one at random from a submitted lot and were classified as defectives or non-defectives.

Obtain the equations for the acceptance and rejection lines for an item-by-item sequential sampling inspection plan in the above case, such that under the plan a lot containing 5 per cent defectives would be accepted in 90 per cent of cases whereas a lot containing 15 per cent defectives would be accepted only in 20 per cent of cases.

Prepare a table of acceptance and rejection numbers for sample sizes from 5 to 50 at intervals of 5. Sketch an approximate O.C. curve of the above plan.

(10+6+9)=25

3 A laboratory carried out an experiment to determine the most suitable chemical solution for giving maximum strength to a yarn of particular quality under identical weaving conditions.

Four solutions S_1, S_2, S_3 and S_4 were available for the experimentation and four looms A, B, C and D were selected at random for the study. Four samples of yarn (four beams) each treated with one of the solutions were obtained and were allocated at random to the four selected looms. The breakage rate on each loom was observed. The above operation was repeated over four periods and the resulting data were arranged in the form of a Latin Square as shown below.

period	Solutions S_1	S_2	S_3	S_4
1	A 5.5	C 2.9	D 9.6	B 6.8
2	C 6.2	D 6.5	B 5.2	A 9.2
3	D 8.9	B 2.9	A 5.8	C 6.6
4	B 6.1	A 5.1	C 2.8	D 9.8

Analyse the above data and recommend the most suitable solution (25)

GROUP C : STATISTICAL METHODS IN GENETICS

(Answer any two questions from this group)

1. Jenkins (1927) obtains three types of families, which give information about the recombination between the genes Y,y and Wx, wx in maize as shown below :

	Phenotypes Y-Wx	Y-wx	y-Wx	y-wx	total
Backcross (coupling)	397	297	289	412	1395
Backcross (repulsion)	78	136	120	80	414
Single backcross (repulsion)	461	161	515	130	1267

[Wx:wx = 3:1 and Y:y = 1:1]

i) Test whether the two single factor ratios are in keeping with the Mendelian expectation.

ii) Obtain the best estimate of recombination fraction.

iii) Test whether the data are homogenous for the recombination fraction thus estimated.

iv) Obtain the standard error of the best estimate of the recombination value. (24)

2. Estimate heritability by doubling the intra-sire regression of daughter (y) on dam (x), from the following data (in units of 10 lbs.) of lactation records of cows (x) and of their daughters (y) :-

Sires A X	A Y	B X	B Y	C X	C Y	D X	D Y	E X	E Y
376	365	484	444	420	432	285	405	427	105
412	447	615	581	388	333	294	43	310	206
366	497	294	664	371	309	524	439	388	242
377	186	497	681	492	138	358	278	227	242
459	435	350	296	484	432	511	304	485	386
285	350	417	392			409	377	358	268
		440	428			351	322	444	275
		55	306			418	425	475	292
						377	365	581	308

What is the regression of y on x within each of the sire progenies separately? What is the pooled intra-sire regression of y on x in the whole set of data? Do the individual regressions differ significantly among the five progenies? (24)

3.(a) 427 blood specimens were collected from 216 children and 211 adults belonging to a particular community, and they were tested by standard methods for blood group antigens. The observed numbers for different phenotypes in two systems of tests (A B O and M-N) are as given below:

A B O Blood groups		M-N Blood groups	
Phenotypes	Observed number	Phenotypes	Observed number
O	122	M M	211
A	97	M N	102
B	163	N N	54
A B	45		
Total	427	Total	427

Determine the gene frequencies of A B O and M-N blood groups, by maximum likelihood solution and gene counting methods respectively. Also calculate the standard errors of these estimates.

Give your comments on the blood group behaviour of the community.

(b) Given the genotypic values for AA, Aa, aa as 1, 0, - 2 respectively, compute the correlation between the genotypic values of full brother-sisters, starting with a population $(p^2 AA + 2pqAa + q^2 aa)$ in which there is random mating.

GROUP D : VITAL STATISTICS AND DEMOGRAPHY

(Answer any two questions from this group)

1.(a) From the following values of some life-table functions obtain

i) the number of years lived by the Cohort within each of the stated age-groups and

ii) also the expectation of life at ages 0, 1, 5, 15, 25 and 35 :

Age-group x to x+n	nq_x	
0 - 1	0.14273	
1 - 5	0.13852	l_0 = 100,000
5 - 15	0.05373	
15 - 25	0.03448	T_{35} = 2,057,009
25 - 35	0.04126	

(b) Write a note on the expectation of life at different ages as obtained in (a). (25)

Please turn over

The following table gives the distributions of the number of children born-to-couples within 5, 6, 7, 8 and 9 years' duration of effective marriage, from the return-marriage time. Find the average number of children per couple per year for each duration. Comment on your findings!

(25)

number of children born	distribution of the number of children born to couples, within effective marriage period duration of				
	5 years	6 years	7 years	8 years	9 years
0	113	65	49	40	31
1	122	117	91	65	52
2	105	109	103	99	79
3	57	76	84	81	88
4	23	35	43	60	62
5	7	12	19	34	43
6		2	5	21	8
7				3	2
Total	427	416	399	384	365

(25)

3.(a) The number of births occurring in a country in a particular year is shown here classified according to age of mother, together with the female population in each age group of the reproductive period.

Age-group	Female population	Number of births to mothers in the age-group
15 - 19	84,708	2,349
20 - 24	70,018	14,647
25 - 29	72,660	16,746
30 - 34	75,924	10,229
35 - 39	75,109	5,267
40 - 44	71,625	1,432

The total population of the country during the year was 2,285,800.

With the above data, determine

i) the crude birth rate,

ii) the general fertility rate,

iii) the age-specific fertility rate and

iv) the total fertility rate.

Assuming that the sex-ratio at birth was 104.6 male births to 100 female births, determine also the gross reproduction rate.

(b) Comment on the values of the different rates obtained in (a).

(25)

GROUP E : EDUCATIONAL AND PSYCHOLOGICAL STATISTICS

(Answer question 4 and any other two questions from the rest of this group)

1. A test was given to 300 individuals with the result that the reliability coefficient was 0.88 and the standard deviation of the obtained scores was 11.50. In a selected sample of 50 individuals from the same group, the standard deviation was 9.60. What is the probable coefficient of reliability for the latter group ?

[Assume the standard error of measurement to be the same in the two groups]

(12)

2. In each of the upper and lower 27 per cent groups on the basis of the distribution of the total score of four items, there are 100 cases. The frequencies of persons answering the individual items correctly in these two groups are given in the table below :

	Items			
	1	2	3	4
Upper 27 %	75	80	95	92
Lower 27 %	40	30	25	30

Find the Upper-Lower Index (ULI) for items 1, 2, 3 and 4 and also the corresponding standard errors.

(12)

3. Test X has a validity coefficient of of 0.65 and reliability 0.75 whereas the validity of test Y is 0.67 and its reliability is 0.95. Each of these tests is a 50-item test.

Which one of them would show greater validity when the test length is increased 4 times ?

(12)

4. The following is a Centroid Factor Loading Matrix (A)

Tests	A_o	B_o	C_o
1	.493	.228	-.718
2	.626	.302	-.328
3	.658	.188	.512
4	.545	.483	.528
5	.681	-.648	-.112
6	.461	-.643	.113

Obtain the rotated factor matrix, after suitable orthogonal rotations of the centroid factor loading matrix (A).

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper VIII: Subjects of Specialisation - I (Theoretical)

Time: 4 hours Full marks:100

(a) Candidates will be required to answer questions from that group only for which they have registered.

(b) Figures in the margin indicate full marks.

Group A: Economic Statistics (Econometrics)

Special Paper - I

(Answer any five questions from this group)

1. Define the concept of Engel elasticity. For calculating an empirical estimate, what type of data you will collect? Give details of the estimation process.

Does Engel elasticity change with changing price structure? [20]

2. Consider a market of two persons and two commodities. The initial allocation is given by the matrix :

$$\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$$

and the utility functions of the two persons are

$$u_1(x_1, x_2) = x_1^{1/2} x_2^{1/2}$$

and $$u_2(x_1, x_2) = x_1^{1/4} x_2^{3/4}$$

What are the indices of utilities of the initial allocation for both persons?

Determine the equilibrium price of commodity 2 in terms of commodity 1.

What will be the final allocation after exchange of the two commodities?

What are the indices of utility after exchange? [20]

3. Discuss the assumptions behind the derivation of individual's 'demand curve'." Market demand curves are obtained by adding together the demand curves of all the individuals in the market. Such curves give the demand-price relationship for those commodities only, whose owners before the price change do not belong to the market".

Discuss the above statement pointing out the assumption that will be violated if the owners of the commodity belong to the market before the price change. What will happen to the market demand in such a case? [20]

4. Explain the concept of 'isoquant' (equal product curve). Derive it from a production function.

The following table gives the output per acre of paddy with various amounts of per acre labour employed.

Output (Kg)	800	900	1,000
Labour (man-days)	40	54	70

4. (contd.)

Assuming constant returns to scale and only two factors of production viz., land and labour; give the coordinates of three points on the isoquant of paddy with output level of 1,000 kg., corresponding to the three output-labour combinations given above.

Which of the three factor combinations the farmer will adopt, if the rent per acre, is more than the wages of 90 man-days? [20]

5.a) Explain the problem of multicollinearity, and develop a technique of dealing with it. Is it always possible to get over this difficulty by suitable estimation procedures?

b) Suppose that two explanatory variables are exactly related according to,

$$Z_{2t} = 3.0 + 0.5Z_{1t} \text{ , and that}$$

$$Y_t = -1.8 + 2.0Z_{1t} - 0.7Z_{2t} + v_t \text{ .}$$

Find an equation for Y_t that is equivalent to the foregoing and that has an arbitrary value α for the coefficient of Z_{2t}. [20]

6. Describe an Input-Output table. How will you obtain national income generated by various sectors from it?

How will you adjust the table for price changes? State if the proportion of national income generated in different sectors will remain invariant to price changes. [20]

7. Discuss critically Mahalanobis model for Indian Planning. Does it ensure consistency or balanced growth? If not, how will you supplement it? [20]

8. In the study on income distribution, it is assumed that the change in the income of an individual is a random proportion of the income of the individual in the last period. Derive the distribution law of income and state its properties. How is this law related to the concentration curve? [20]

Group B: Techno-Commercial Statistics

(Statistical Quality Control) - Special Paper - I

(Answer any four questions from this group)

1.a) What is a control chart? State evidences of lack of control in a manufacturing process as provided by a control chart.

b) Distinguish between 3σ limits and probability limits as used on a control chart.

c) Describe the advantages of maintaining a group control chart.

d) Indicate situations to use the following:-
(i) $\bar{X}$ - R charts, (ii) chart for the number of defectives, (iii) modified control chart and (iv) chart for moving averages. [2+4+4+5+10]=25

(a) Explain how would you develop standard values for $(\bar{X}-R)$ chart from the analysis of historical data.

(b) Explain what is meant by process capability and discuss how it can be estimated.

(c) If the nominal standard (norm) for some quality characteristic (x) of a product be 180 and the specifications be +2 per cent and -8 per cent of the nominal, obtain the percentage of items within specifications with $\bar{X} = 176$ and $\sigma = \frac{\bar{R}}{d_2} = 4$. It is given that

$$\int_0^{1.9} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = .47128 \text{ and } \int_0^{2.6} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = .4$$

[12+8+5]=2

(a) Explain the nature of quality protection afforded to the custer by the various types of lot - by - lot sampling inspection plans, by attributes.

(b) Explain the method of constructing a single sampling AOQL plans used in Dodge-Romig tables. State necessary assumptions. [10+10+5]=2

(a) What features of a sampling inspection plan are brought out its operating characteristic curve?

(b) Find out the expected total cost of operating a double sampling acceptance-rejection plan, taking into account the cost of sampling, cost of inspection and the cost of accepting defective item. Deduce necessary expressions. [10+15]= 2

(a) Describe briefly a sequential sampling inspection plan by attributes, given AQL and LTPD; producer's risk and consumer risk. Show how it can be carried out graphically.

(b) Discuss the salient features of Bowker-Goode sampling inspection plans for variables. [15+10]= 2

Write notes on any four of the following:

i normal, reduced and tightened inspection,

ii use of rotatable designs in industrial experimentation,

iii drawbacks of stop-watch time study as a method for work measurement,

iv difficulties in sampling from bulk material,

v selection of rational sub-groups for purposes of maintaining a control over current production,

vi expressions for c_2, B_3 and B_4 as used in control charts. [1

Group C: Biometric Methods (I)
(No candidate available)

Please Turn Over

Group D: Design and Analysis of Experiments(I)
(No candidate available)

Group E: Sample Surveys (Theoretical Aspects)
- Special Paper - I

(Answer any four questions from this group)

1.(a) Consider simple random sampling (srs) without replacement of n units from a population of N units. Prove that the probability of selecting unit i at r^{th} draw and unit j at s^{th} draw equals $1/[N(N-1)]$ where $i \neq j = 1,\ldots, N$ and $r \neq s = 1,\ldots, n$.

(b) Using (a) or otherwise prove that the correlation coefficient between y'_r and y'_s is $-1/(N-1)$ where y'_r denotes the value of the unit drawn at r^{th} draw.

(c) Using (b) or otherwise, derive the variance of the sample mean $\bar{y} = n^{-1} \Sigma y'_r$. [7+8+10]= 25

2.(a) For simple random sampling (srs) of n units with replacement from a population of N units, denote $a_i = 1$ if i^{th} unit is at least once in the sample and zero otherwise. Let n denote the number of distinct units in the sample. Express n in terms of the a_i and find the expected value of n, E (n). Prove that $E(n) < n$.

(b) Suppose $n = 3$ in (a). Derive the probabilities that the sample contains 1, 2 and 3 distinct units. Noting that for a fixed n we have (srs without replacement, prove that the variance of the unweighted mean of the values of the distinct units in the sample is equal to

$$\frac{(2N-1)(N-1)}{6N^2} S^2$$

where $(N-1)S^2 = N\sigma^2$ and σ^2 is the population variance.

(c) Consider srs without replacement of n units and assume that the finite population is a random sample from an infinite super-population which is normally distributed. Prove that:

$$t = \frac{\sqrt{n}\,(\bar{y} - \bar{Y})}{s\,\sqrt{(1-n/N)}}$$

is exactly distributed as Student's t with (n-1) degrees of freedom, where $\bar{y}$ and $\bar{Y}$ are the sample and population means respectively and

$$s^2 = (n-1)^{-1} \Sigma (y_i - \bar{y})^2.$$

[Hint: $\bar{Y}$ is a linear combination of $\bar{y}$ and the mean of the remaining $(N - n)$ units.] [7+10+8]= 25

3.(a) Let X_i denote the size of i^{th} unit $(i=1,\ldots,N)$ in the population. Describe Lahiri's method of selecting one unit with probability proportional to size (pps).

3.(b) Prove that the probability of selection of i^{th} unit for Lahiri's method, equals X_i / X where X = population total of the X_i.

(c) For unequal probability sampling of n units without replacement $(n > 1)$, derive the variance of the Howitz-Thompson estimator

$$\hat{Y}_{HT} = \Sigma y_i / \pi_i,$$

where π_i = probability of inclusion of the i^{th} unit in the sample.

(d) Suppose the first unit is selected with pps of the X_i and then (n-1) units with srs without replacement, from the remaining (N-1) units. Derive π_i and show that the probability of selection of a sample of n units is proportional to the total of the sizes of the units in the sample. [3+7+7+8]= 25

4.(a) Consider the difference estimator $\bar{y}_d = \bar{y} + b_o (\bar{X} - \bar{x})$ for srs without replacement, where $\bar{y}$ and $\bar{x}$ are sample means, $\bar{X}$ is population mean and b_o is a pre-assigned constant. Derive the value of b_o which minimises the variance of $\bar{y}_d$. You may assume the formulae for the variances of $\bar{y}$ and $\bar{x}$ and covariance of $\bar{y}$ and $\bar{x}$.

(b) Prove that if $[V(\bar{y}_d) - V'(\bar{y}_d)] / V'(\bar{y}_d)$ is to be less than α, then

$$\left| \frac{b_o}{\beta} - 1 \right| < [\alpha(1-\rho^2)/\rho^2]^{1/2}$$

where ρ and β are the population correlation and regression coefficients respectively and $V'(\bar{y}_d)$ is the variance with $b_o = \beta$.

(c) Derive an exact upper bound to the bias of the classical ratio estimator relative to its standard error.

(d) Derive the Hartley-Ross unbiased ratio estimator for srs without replacement. [3+8+7+7]=25

5.(a) Suppose the population consists of N primaries with M secondaries in each primary. Obtain an unbiased estimator of the population mean and derive its variance, assuming that n primaries and m secondaries within each selected primary, are sampled with srs without replacement.

(b) Suppose there are two strata. From stratum 1 a srs sample of n_1 units is selected without replacement from the N_1 units in that stratum. In stratum 2 there are N_2 primaries with M_2 secondaries in each primary. A srs sample of n_2 primaries and m_2 secondaries from each selected primary is drawn without replacement. Suppose the cost function is $C = n_1 c_1 + n_2 c_2 + n_2 m_2 c_3$ where c_1 = cost per unit in stratum 1, c_2 = cost per primary and c_3 = cost per secondary in stratum 2.

Derive...

5.(b) contd.

Derive an unbiased estimator of the population mean and then derive the optimum value of n_2, which minimises the variance of the estimator for fixed C.

(c) Consider a four stage design in which ($n > 1$) primaries are drawn with srs with replacement. Each time a primary is selected, a sample of secondaries is selected independently with unequal probabilities without replacement. Within each selected secondary, a sample of third stage units is selected with srs without replacement and from each selected third-stage unit, a systematic sample of fourth stage units is selected circular systematically. Denote by

$$\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$$

the unbiased estimators of the n primary totals. Obtain an unbiased estimator of the population total and derive an estimator of its variance in terms of

$$\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n. \qquad [9+10+6]=25$$

6.(a) Questionnaires were mailed to a srs sample of n persons and n_1 of them responded. From the $n_2 = n - n_1$ non-respondents a sub-sample of $r_2 = n_2/k$ persons is selected with srs again and the values are obtained by direct interview, k being fixed. Derive an unbiased estimator of the population mean. Derive its variance.

(b) If the cost function is $C = c_0 n + c_1 n_1 + c_2 r_2$ where c_0 = cost per mailing questionnaire, c_1 = cost of processing per questionnaire and c_2 = cost per direct interview, derive the optimum values of n and k which minimise the expected cost (EC), subject to a fixed variance V_0.

(c) Describe briefly the merits of interpenetrating subsamples. [10+10+5]=25

Group F: Techniques of Computation

(Numerical Analysis) - Special Paper - I

(Answer five questions from part A and one question from part B)

PART A

1.(a) The function $f(x) = \log_{10} x$ is tabulated correct to 6 decimal places for $x = .5$ to $x = .6$ at an interval of .001. What is the error in $\log_{10} x$ obtained by linear interpolation for $.5 \leq x \leq .6$ given correct to six places of decimal.

(b) If linear interpolation in (a) is to give $\log_{10} x$ correct to six places of decimal for any given x in (.5, .6) without error, what should be the interval of tabulation?

(c) Suggest a method of condensing the table in (b). [6+5+5]=16

2. By integrating the appropriate central difference approximation formula for $f(x)$, obtain the quadrature formula

$$\frac{1}{h}\int_{x_0}^{x_1} f(x)dx = \frac{1}{2}(f_0 + f_1) - \frac{1}{12}(\mu\delta f_1 - \mu\delta f_0)$$

$$+ \frac{11}{720}(\mu\delta^3 f_1 - \mu\delta^3 f_0) - \frac{191}{60480}(\mu\delta^5 f_1 - \mu\delta^5 f_0) + \ldots$$

where

$$f_i = f(x_i) \qquad i = 0,1$$

$\delta^r f$ is the r^{th} central difference, and

$$\mu\delta^{r+1} f_j = \frac{1}{2}\left(\delta^r f_{j+\frac{1}{2}} + \delta^r f_{j-\frac{1}{2}}\right) \qquad [16]$$

3. Suppose that the equation $x^2 + a_1 x + a_2 = 0$ possesses real roots α and β. Show that the iteration $z_{k+1} = -(a_1 z_k + a_2)/z_k$ is stable near $x = \alpha$ if $|\alpha| > |\beta|$; the iteration $z_{k+1} = -a_2/(z_k + a_1)$ is stable near $x = \alpha$ if $|\alpha| < |\beta|$; and the iteration $z_{k+1} = -(z_k^2 + a_2)/a_1$ is stable near $x = \alpha$ if $2|\alpha| < |\alpha + \beta|$. [5+5+6]=16

4. Using Chebychev polynomials, obtain an approximation for $\cos x$ of the form

$$A_0 + A_2 x^2 + A_4 x^4$$

with an error smaller than 5×10^{-5} over $(-1, 1)$. [16]

5. Show that the eigen values λ_i, $i = 1,2,\ldots, n$, of a real $n \times n$ matrix $A = ((a_{ij}))$ satisfy:

i) $$|\lambda_i| \leq \min\left\{\max_i \sum_{k=1}^{n} |a_{ik}|, \max_k \sum_{i=1}^{n} |a_{ik}|\right\}$$

ii) $$|\lambda_i| \leq \left[\left(\max_i \sum_{k=1}^{n} |a_{ik} + a_{ki}|\right)^2 + \left(\max_i \sum |a_{ik} - a_{ki}|\right)^2\right]^{\frac{1}{2}} / 2.$$

[6+10]=16

6. $A_{n \times n}$ is a non-singular matrix which is partitioned as under:

$$A_{n \times n} = \begin{bmatrix} A_{11} & \vdots & A_{12} \\ \cdots & \vdots & \cdots \\ A_{21} & \vdots & A_{22} \end{bmatrix}$$

and

$$A^{-1} = \begin{bmatrix} A^{11} & A^{12} \\ A^{21} & A^{22} \end{bmatrix}$$

such that $A_{11}(p \times p)$ is non-singular and A^{11} is also of dimension $(p \times p)$.

6. (contd.)

i) Show that $A^{22}_{(q \times q)}$ is non-singular (where q=n-p).

ii) Obtain the inverse of A in terms of the partitioned matrices, suitable for computation. [8+8]=16

7 : PART B

7.(a) If $y = f(x)$ and $f'(x) \neq 0$ for $x_0 < x < x_1$, show that the error of the linear inverse interpolation based on corresponding values (x_0, y_0) and (x_1, y_1) is given by

$$-(y-y_0)(y-y_1)\frac{f''(\xi)}{2[f'(\xi)]^3}$$

where $x_0 < \xi < x_1$ if $f''(x)/[f'(x)]^3$ exists and is continuous in that interval.

(b) Show also that the magnitude of the error is limited by

$$\frac{h^2}{8}\left(\frac{M_1}{m_1}\right)^2 \frac{M_2}{m} \text{ for } x_0 \leq x \leq x_1$$

where $m_1 \leq |f'(x)| \leq M_1$,

$|f''(x)| \leq M_2$

and $|x-x_0| = h$. [14+6]=20

8. $f(x,y)$ is continuous in the region R defined by $|x-x_0| < A$ and $|y-y_0| < B$ and $|f(x_0, y_0)| < M$. In addition $f(x,y)$ satisfies the Lipschitz condition $|f(x,y_1) - f(x,y_2)| < L|y_1 - y_2|$. Show that the procedure of successive substitution given by

$$y_{n+1} = y_0 + \int_{x_0}^{x_0+h} f(x, y_n)dx$$

converges to the solution at $x = x_0+h$ of

$y' = f(x, y)$ with initial condition, $y = y_0$ when $x = x_0$

for a proper choice of h. A, B, M and L are positive constants. [20]

Group G: Statistical Inference (Statistical Theory)
- Special Paper - I -
(Answer any five questions from this group)

1.(a) State and prove the Rao-Blackwell theorem in the theory of estimation.

(b) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a Poisson population with parameter λ. Define

$$g(X_1) = 0 \text{ if } X_1 = 0 \text{ and } g(X_1) = 1 \text{ if } X_1 \geq 1.$$

i) Show that $g(X_1)$ is an unbiased estimator of $1 - e^{-\lambda}$.

ii) Starting with $g(X_1)$ obtain a better estimator of $1 - e^{-\lambda}$ by using the Rao-Blackwell theorem.

iii) State, with reasons, whether the estimate you arrive at in (ii) above will be same whatever the unbiased estimator g you start with. [2+4+3+7+4]=20

2.(a) Explain the terms (i) error of the first kind (ii) randomised test and (iii) power-function.

b) $X_1, X_2, \ldots, X_n$ are n independent observations on a random variable X, with density function $\frac{1}{\theta_1} e^{-x/\theta_1}$ and $Y_1, \ldots, Y_n$ on Y, with density function $\frac{1}{\theta_2} e^{-x/\theta_2}$. The X - observations and Y - observations are themselves independent.

Derive the likelihood ratio test for $H_0: \theta_1 = \theta_2$ and show that the test statistic is a function of $\bar{X} / \bar{Y}$ where $\bar{X} = \frac{1}{n} \sum_1^n X_k$; $\bar{Y} = \frac{1}{n} \sum_1^n Y_k$.

What is the distribution of $\bar{X} / \bar{Y}$ (i) under H_0 and (ii) when $\theta_1 = 2\theta_2$? [2+2+2+6+8+3+3]=20

3.(a) Define (i) similar tests and (ii) tests having Neyman structure.

(b) Examine if the usual t-test for testing $\theta = 0$ against $\theta \neq 0$ in a normal population $N(\theta, \sigma^2)$, has Neyman structure.

(c) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the uniform distribution on (θ_1, θ_2). Explain in detail, how you would construct a Uniformly Most Powerful (UMP) unbiased test for testing $H_0: \theta_1 \leq 0$ against $H_1: \theta_1 > 0$.

[2+2+6+10]=20

Please Turn Over

4.(a) Independent samples, each of size n, are taken from populations with frequency functions

$$\frac{1}{\theta} e^{-x/\theta} \quad \text{and} \quad \theta e^{-\theta x}.$$

Show that (X, Y), where $X = \Sigma x_i$, $Y = \Sigma y_i$, is sufficient for θ. Show further that (X, Y) is not complete, by considering the function $XY - 1$.

(b) Use the joint distribution of two order statistics to obtain confidence intervals for any quartile of a continuous distribution. [7+4+9]=20

5.(a) Explain how the problem of finding the best invariant estimator of a scale parameter θ can be reduced to a similar problem on a location parameter.

(b) Let $(X_1, X_2, \ldots, X_n)$ be a random sample from the uniform distribution over $(0, \theta)$. Find the best invariant estimator of θ if the loss function is

i) $L(\theta, a) = \left(\frac{a}{\theta} - 1\right)^2$

ii) $L(\theta, a) = \begin{cases} 0 & \text{if } \frac{1}{c} \leq \frac{a}{\theta} \leq c \\ 1 & \text{otherwise} \end{cases}$

[5+7+8]=20

6.(a) If $\hat{\theta}$ is the maximum likelihood estimator of θ, prove that $\tau(\hat{\theta})$ will be the maximum likelihood estimator of $\tau(\theta)$.

(b) Consider a power series distribution $P(X=k) = a_k \frac{c^k}{f(\theta)}$, $k = 0,1,2,\ldots$ from which is available a set $X_1, X_2, \ldots, X_n$ of n independent observations. Write $\bar{X} = \frac{1}{n}(X_1 + \ldots + X_n)$. Show that the maximum likelihood estimator of θ is a root of the equation

$$\bar{x} = \frac{\theta f'(\theta)}{f(\theta)}.$$

If X is binomially distributed as $B(N,p)$, estimate p by solving the above equation explicitly. [7+6+7]=20

Group II: Probability Theory - (I)
(No candidate available)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1968

Paper IX: Subjects of Specialisation - II (Theoretical)

Time: 4 hours Full marks: 100

(a) Candidates are required to answer questions from that group only for which they have registered their options.

(b) Figures in the margin indicate full marks.

Group A: Economic Statistics - (Indian Economics and Economics of Planning) - Special Paper II

Section I - Indian Economics

(Answer any three questions from this section)

1. Discuss the pattern of institutional reforms that you consider necessary to bring about lasting agricultural progress in India. [16]

2. Comment on the trend of India's national income growth and its distribution by industrial origin since 1951. [16]

3. Give a critical appraisal of the rôle of the public sector in the industrialization of India after independence. [16]

4. Analyse the bearings of the present state of Indian agriculture on the country's recent economic crisis of recession amidst inflation. [16]

5. Examine the reasons for the chronic tendency of India's exports to be less than her imports. Suggest remedies for this problem. [16]

6. Discuss how monopolistic developments in the structure of organised private business in India, affect the task of economic planning within the framework of a mixed economy. [16]

Section II - Economics of Planning

(Answer any three questions from this section)

7. Explain the procedure of construction of inter-industry tables and demonstrate the equivalence of the different concepts of national income, with the help of inter-industry accounts. [16]

8. Explain how does the Mahalanobis two-sector model demonstrate the priority of investment goods industries, for achieving long-run economic growth. [16]

9. Examine how far internal voluntary savings proved to be adequate for fulfilling the investment targets of the second and third Five Year Plans in India. Assess the rôle and consequences of deficit financing and foreign loans in this connection. [16]

10. Give your suggestions for the reformulation of the broad strategy of economic planning in India, in the light of the experience of the last two Five Year Plans. [16]

11. Evaluate the experience of the Indian economy during and after the Third Five Year Plan with reference to (a) price movements, (b) industrial production, (c) agricultural production and (d) problem of unemployment. [16]

(X40) Please Turn Over

12. Review your understanding of any two of the following:

(a) Vicious circles of underdevelopment

(b) role of agricultural surplus in economic development

(c) warranted rate of growth

(d) indirect labour cost of a commodity in the static Leontief system. [8+8]=[16]

Neatness (Sections I and II) - [4]

Group B: Techno-Commercial Statistics - Special Paper II

Group B: Techno-Commercial Statistics (II)
Sec.I: Operations Research
(No candidate available)

Section I: (Alternative) - Elements of Book-keeping and Accountancy

Full marks: 70

N.B. (a) Use separate answer book for this section.
(b) Answer Q. No.1 and any three questions from the rest of this section.

1. From the following Trial Balance and notes, make out (a)
(a) Trading and Profit and Loss Account for the year ended 31st March, 1968 and (b)
(b) a Balance Sheet as on that date: (c)

	Dr. Rs.	Cr. Rs.
Material consumption -	1,80,000	
Wages -	50,000	
Salaries -	25,000	
Sales and Returns -	3,000	4,30,000
Closing Stock of Materials -	20,000	
Factory Expenses -	30,000	
Opening Stock of Finished goods -	15,000	
Capital and Drawings:		
A -	8,000	50,000
B -	5,000	40,000
Rent -	6,000	
Plant and Machinery -	1,60,000	
Debtors and Creditors -	40,000	30,000
Bad Debts -	3,000	
Provision for Bad Debts -		8,500
Furniture -	4,000	
Insurance -	3,000	
Postage -	500	
Rates and Taxes -	300	
Cash -	1,000	
Bank -	4,000	
Discount Allowed -	700	
	5,58,500	5,58,500

Notes:
(1) Opening Stock of Materials Rs.25,000.
(2) Closing Stock and Finished Goods Rs.20,000.
(3) Provide for 10 per cent depreciation on all Fixed Assets.
(4) Provision for Bad Debts to be maintained at 5 per cent on Sundry Debtors.
(5) Outstanding Salary Rs.1,000.

Please Turn Over

1.(contd.)

Notes:

(6) Provide 6 per cent interest on capitals.
(7) Profit is shared equally by A and B. [10+7+8]=[25]

2.(a) Define 'transaction' with 3 illustrations.

(b) Write a short note on the 'Rules of Journalising', as applicable to different types of Accounts.

(c) What is 'Journal Proper'? What are its uses?[4+5+(3 X2)]=[15]

3. Make out a Cash Book with imaginary figures for:

(a) Opening Cash and Bank balances.
(b) 2 withdrawals from Bank.
(c) 2 deposits of cash from cash balance.
(d) Payments - one cash and one cheque at least.
(e) Receipts - one cash and one cheque at least. [15]

4. Pass entries to rectify the following errors:

(a) Repairs to furniture Rs.500, debited to Furniture Account.
(b) Rs.538 received from Mr. A credited to him as Rs.358.
(c) Rs.500 paid to Mr. X, credited to him.
(d) Rs.362 received from Mr. B debited to A's Account.
(e) Rs.231 paid to Mr. C credited to him as Rs.132. [3 X 5]=[15]

5.(a) What is columnar Petty Cash Book ?

(b) Discuss in details, the meaning and Operation of Imprest System of Petty Cash.

(c) Explain the difference between Trial Balance and Balance Sheet. [4+6+5]=[15]

6. Make out Ledger Accounts for the following transactions:-

(a) Rent paid Rs.5,500 and Rent outstanding Rs.500. Last year's outstanding was Rs.1,000.

(b) Balance of Furniture Account is Rs.4,000 Furniture was purchased for Rs.1000. Provide depreciation 10 per cent on opening balance.

(c) Salary Rs.13,000 paid. Salary paid in advance last year was Rs.500. Salary outstanding this year is Rs.1,000. [5X3]=[15]

Section II: Statistical Methods in Business -(30 marks)

(i) Use a separate answer book for this section.
(ii) Answer any two questions from this section.

1.(a) How would you estimate the demand for television set in a town where a television station is going to be set up soon? How would you estimate it for the whole country during the next ten years?

(b) Give some business forecasting techniques. [7+8]=[15]

2.(a) Explain the use of statistical methods (i) job evaluation and personnel selection (ii) evaluating the effectiveness of bonus and incentive schemes.

(b) Explain the use of sampling methods in accountancy and auditing. [4+4+7]=[15]

3. A firm can advertise its goods through (i) daily newspapers, (ii) weeklies, (iii) monthlies, (iv) radio (v) post. It wants to go on improving its advertising policy steadily in the light of the results obtained from sample surveys. Explain how you would advise the firm to plan its advertisement campaigns. [15]

Group C: Biometric Methods
(No candidate available

Group D: Design and Analysis of Experiments - Special Paper II
(Combinatorial Aspects).

(a) Answer any five questions from this group.

1.(i) Show that the ring M. of integers modulo n is a field, if and only if n, is a prime number.

(ii) If a field F is finite and has n elements, show that n is of the form p^n where p is the prime characteristic of the field F and n is a positive integer.

(iii) Show that for GF(3^2), x^2+1, x^2+x+2 and x^2+2x+2 may each be taken as a minimum function. Determine, in general, the number of polynomials of degree n with coefficients from GF(p) and irreducible in the same field. [6+6+8]=[20]

2. Let $p_1^{n_1} p_2^{n_2} \ldots p_u^{n_u}$ be the canonical prime power decomposition of v and N(v) = maximum number of mutually orthogonal latin squares of order v, and where

$n(v) = \text{Min}\ (p_1^{n_1}, p_2^{n_2}, \ldots p_u^{n_u}) - 1$. Show that

(i) $n(v) \leq N(v) \leq v-1$

(ii) in general, $n(v) \neq N(v)$. [10+10]=[20]

3. State necessary conditions known to you about the existence of symmetrically balanced incomplete block (BIBD) designs. Hence establish the impossibility of the following BIBD: (22, 22, 7, 7, 2), (43, 43, 15, 15, 5), (77, 77, 20, 20, 5). [8+4+4+4]=[20]

4. Let N_1 be the incidence matrix of the BIBD: $(v_1, b_1, r_1, k_1, \lambda_1)$ where $b_1 = 4(r_1-\lambda_1)$, $i = 1,2$, and let M_1 be obtained from N_1 by changing the null elements into -1. Show that the design N obtained from the Kronecker product $M_1 \times M_2$, by changing the elements -1 to 0 is a BIBD. Obtain the parameters of this new BIBD. [15+5]=[20]

5. Let N be the incidence matrix of a PBIBD and let $B_t = (a_{ijt})$, where $a_{ijt} = 1$ if the treatment pair (i,j) is a t-associate and zero otherwise, $i,j = 1,2,\ldots, v$; $t = 1,2,\ldots, m$. Establish that :

(i) $NN' = r I_v + \sum_{t=1}^{m} \lambda_t B_t$

(ii) the multiplicity of the roots of NN' is independent of $r_1 \lambda_1, \lambda_2, \ldots, \lambda_m$. [10+10]=[20]

Please Turn Over

6. Let N be the incidence matrix of a binary equi-replicate design of constant block size k, and x the number of common treatments between two blocks of the design. Show that x satisfies the inequalities

$$\text{Max}\,(0,\ 2k-v,\ k-\mu) \leq x \leq \text{Min}\left\{k,\ -k+\mu+\frac{2(rk-v)}{b}\right\},$$

where r, k, v and b have their usual meanings and μ is the largest characteristic root of NN' smaller than rk.

Show that if $b = 2(v-1)$, a balanced incomplete block design with parameters v, b, r, k and λ can never be resolvable if k is odd. [10+10]=[20]

7. In a symmetrical factorial experiment with n factors each at s levels, if two treatments are called i-associates ($i = 1,2,\ldots, n$) when the levels of exactly n-i factors in one treatment are the same as the levels of the corresponding factors in the other treatment, show that such an association scheme satisfies the requirements for the association scheme of a PBIBD. Obtain n_i and p_{ij} for this association scheme. [20]

8. Describe a method of construction of a $\frac{1}{s^k}$ replicate of a s^n design and explain how in such experiments confounding can be resorted to, for effective elimination of soil heterogeneity.

How can a set of (s-1) mutually orthogonal Latin Squares of side s, be used to construct an $\frac{1}{s^{s-1}}$ replicate of an s^{s+1} design? Write a critical note on the merits and demerits of fractional replicates obtained this way? [7+3+7+3]=[20]

Group E: Sample Surveys ... - Organizational Aspects

Special Paper II

(Answer any five questions from this group)

1. Describe various methods of data collection in a sample survey. Discuss their merits and demerits. Mention situations of preferring one particular method to others. [10+5+5]=[20]

2. What are the factors that should be taken into consideration while suggesting the qualification of various categories of investigating and supervisory staff in a sample survey? Discuss the advantages and disadvantages of employing whole-time staff, part-time staff and existing administrative staff for conducting the field work of a large scale sample survey. [10+10]=[20]

3. What are cost and variance functions with reference to a sample survey? What type of data do you require for building up these functions and how do you obtain them? Describe, with a suitable example, the method of determining the sample size which is likely to give the estimate with maximum precision for a given cost. [5+5+10]=[20]

4. Explain clearly the concept of non-sampling errors in surveys. What are the various reasons for non-sampling errors? What steps will you suggest to control them? [4+8+8]=[20]

5.(a) Describe various stages of processing of data collected in a survey.

(b) Describe methods of controlling numerical errors in large scale statistical calculations.

(c) Give some examples to show that defects in sampling frame may sometimes be compensated by suitably adjusting the results. [8+6+6]=[20]

6. It is proposed to organise a post-enumeration check survey for checking the quality of data planned to be collected in 1971 population census of India.

(i) What should be the sampling unit and sampling frame?

(ii) What stratification would you suggest?

(iii) What are the items on which you will like to collect data?

[Note: The villagewise data of the last population census are available.] [7+7+6]=[20]

7. Write short notes any three of the following:-

(a) Master sample

(b) Budget and cost control

(c) Mechanical tabulation

(d) Scrutiny of final report. [20]

Group F: Techniques of Computation - Special Paper II Practical

Time: 5 hours Full marks: 100

(a) Answer any five questions from this group.
(b) Figures in the margin indicate full marks.
(c) Use of Calculating Machines is permitted.

1.(a) Find the number of terms required in the expansion

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \cdots$$

to compute 'e' correct to 8 places of decimals.

(b) Van der Waals' equation of state for a real gas is expressed as

$$V^3 - (b + \frac{RT}{p})V^2 + \frac{a}{p}V - \frac{ab}{p} = 0$$

where a, b and R are constants. The equation becomes a cubic in V (volume) for constant values of p (pressure) and T (temperature). Assume that for a certain gas there exists $p = p_c$ and $T = T_c$ such that the cubic equation in V has all roots real and equal. Let V_c denote the value of the root. Find the accuracy required in the determination of the constants a, b and R so that V_c, p_c and T_c may be determined to within 0.05 percent of their respective true values.

Please Turn Over

2. Using the following table of values of the incomplete normal probability integral

$$F(x) = \int_{x}^{\infty} (2\pi)^{-\frac{1}{2}} \exp(-t^2/2)dt$$

find by interpolation the values of

$$I(x) = \int_{-x}^{x} (2\pi)^{-\frac{1}{2}} \exp(-t^2/2)dt$$

for $x = 0.01, 0.675, 1.05$

x	F(x)
0.0	0.50000
0.1	0.46017
0.2	0.42074
0.3	0.38209
0.4	0.34458
0.5	0.30854
0.6	0.27425
0.7	0.24196
0.8	0.21186
0.9	0.18406
1.0	0.15866
1.1	0.13567

3. Find by Euler's summation formula, or otherwise the sum

$$S = \frac{1}{400} + \frac{1}{402} + \frac{1}{404} + \ldots + \frac{1}{600} .$$

4. Find the real root, correct to five places of decimal of the following equation :

$$x - \frac{x^3}{3} + \frac{x^5}{10} - \frac{x^7}{42} + \frac{x^9}{216} - \frac{x^{11}}{1320} = 0.4431135$$

5. Tabulate the numerical solution of

$$\frac{dy}{dx} = \text{Sin } x + \text{Cos } y \qquad x = 30^\circ,\ y = 45^\circ$$

for $x = 35^\circ\ (5^\circ)\ 60^\circ$.

6. Find the largest latent root and the associated normalised latent vector of the symmetric matrix :

$$A = \begin{vmatrix} 2.296 & 0.067 & -0.567 & 0.709 \\ & 0.499 & 0.113 & 0.311 \\ & & 0.417 & 0.070 \\ & & & 1.472 \end{vmatrix}$$

7. Solve the equations:

$$15.129x_1 + 23.860x_2 + 1.793x_3 + 0.998x_4 = 4.583 \qquad (i)$$

$$23.860x_1 + 54.756x_2 + 3.635x_3 + 3.511x_4 = 10.091 \qquad (ii)$$

$$1.793x_1 + 3.635x_2 + 18.225x_3 + 21.122x_4 = 29.156 \qquad (iii)$$

$$0.998x_1 + 3.511x_2 + 21.122x_3 + 60.516x_4 = 60.173 \qquad (iv)$$

Group G : Statistical Inference (II)
(No candidate available)

Group H : Probability Theory (II)
(No candidate available)

INDIAN STATISTICAL INSTITUTE

Statistician's Diploma Examination - November 1958.

Paper X - Subjects of Specialisation - III (Practical).

Time: 5 hours Full marks: 100

i) Candidates are required to answer questions from that group only for which they have registered their options.

ii) Figures in the margin indicate full marks.

iii) Use of calculating machine is permitted.

GROUP A: ECONOMIC STATISTICS (Special Paper III -Practical)

(Answer any three questions from this group)

1. The expansion programme for the production of electric power in a country envisages an annual increase in output by 7200 MWh. The new power stations built for this purpose should have a total peak capacity of 2307 MW with a guaranteed total minimum capacity of 1692 MW. The plan for the construction of power stations takes into account five types of power stations: (I) thermal power stations, (II) hydroelectric stations with reservoirs, (III) hydroelectric stations on rivers, (IV) power stations with sluice installations, (V) power stations operated by the ebb and tide of the ocean.

The following table presents the principal technical data per unit of guaranteed capacity of these types of power stations.

	TYPE OF POWER STATIONS				
	I	II	III	IV	V
Guaranteed Capacity (M.W.)	1	1	1	1	1
Peak Capacity (M.W.)	1.15	1.20	1.10	3	2.15
Yearly Output (M.W.h)	7	1.30	1.20	7.35	5.45
Building Costs - investments (millions of National Currency)	97	130	420	310	313
Yearly Operating Costs (Million of National Currency)	136	101	56	140	79

Derive the optimum solution if the joint construction and operating costs, discounted at 8%, have to be minimised.

Compare the derived results with the requirements of the plan.

(32)

Please turn over

2. Agricultural output in year t (Q_t) of a country A is related to its land input (A_t) and labour input (L_t) through the following production function relation:

$Q_t = (A_t)^a \ (L_t)^b \ (10)^c + u_t$ where a, b, c are parameters and u_t is a random disturbance, which has the following properties:

i) $E(u_t) = 0$ for all t

ii) $E(u^2_t) = \sigma^2$ for all t

iii) $E(u_t u_{t'}) = 0$ for $t \neq t'$

Assuming that A_t and L_t are non-stochastic variables, derive the least linear unbiased estimates of a, b, c from the data of table below:

t	A_t (million acres)	L_t (million man-years)	Q_t (million tons)
0	62.1	21.8	5.35
1	64.3	22.2	5.56
2	66.5	22.5	5.68
3	71.0	22.8	5.93
4	70.5	23.2	5.98
5	72.5	24.0	6.31
6	75.4	24.5	5.93
7	76.1	25.0	6.31
8	77.9	25.5	6.67
9	82.8	26.1	6.43
10	83.3	26.7	6.85

Derive an unbiased estimate of the variance-covariance matrix of your estimates of a, b and c.

Test the hypothesis of constant returns to scale. (14)

3. You are given time series data (relating to demand for textiles in a certain country for the period 1923-1939) on the following variables.

y = logarithm of consumption per head

x_1 = logrithm of real income per head

x_2 = logarithm of the deflated price of the commodity

The following matrix gives the sums (in the last column), sum of squares and sum of products of these variables:

	y	x_1	x_2	1
y	76.658975	72.596420	67.441918	36.0763
x_1		68.841788	64.064566	34.2078
x_2			59.759499	31.8339
1				17

Calculate the income and price elasticities of demand and their standard errors. (14)

4. The following table gives the per capita monthly expenditure on all items and on sugar, in urban areas by per capita expenditure classes.

National Sample Survey, 12th Round - (March-August 1957)

monthly per capita expenditure classes (in rupees)	estimated percentage of population	monthly per capita expenditure (in rupees)	
		all items	sugar only
0- 8	14.71	6.49	0.10
8-11	18.36	9.49	0.22
11-13	11.85	12.02	0.29
13-15	10.76	14.00	0.41
15-18	12.92	16.50	0.62
18-21	8.04	19.53	0.72
21-24	6.02	22.57	0.73
24-28	6.41	25.79	0.92
28-34	3.69	30.84	1.09
34-43	2.84	38.25	1.26
43-55	1.70	49.29	1.49
55-above	1.80	105.34	1.77

Draw the Lorenz curve for total expenditure, concentration curve for sugar and relative concentration curve for sugar.

Either, fit a constant elasticity Engel curve for the consumption of sugar:

Or, fit a function of the form $Y = \begin{cases} \frac{\alpha (y - \gamma)}{(x + \beta)} & \text{for } X \geq \gamma \\ 0 & \text{for } X < \gamma \end{cases}$ for the consumption of sugar.

[Assume an arbitrary value for γ (gamma).]

(32)

5. Below is given the transactions matrix for Poland for the year 1956.

(a) Prepare a small note commenting on the interesting features of the inter-industrial and other transactions of the Polish economy. Illustrate your answer by percentages, etc. derived from the table.

(b) Obtain a 3x3 inter-industry transactions table by simply aggregating sectors 2 and 3 ; 1, 4 and 7 ; and 5 and 6. Work out the corresponding input-output coefficient matrix. (32)

Neatness (4)

Please turn over

Input-output table : Poland 1956

(Figures in thousand of millions of Zlotys)

Input from industries	Output to industries								Net output						Gross output
	Industry and Handicrafts	Agriculture	Forestry	Building	Transport and Communication	Trade	Remaining material output	Total	Exports	Individual Consumption	Collective Consumption	Investments and Capital Repairs	Increase in Stocks and Reserves	Losses and Sums not Balanced	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	
1. Industry and handicrafts	125,8	9,0	0,1	13,4	4,1	3,2	0,4	158,3	25,9	106,0	3,4	14,7	12,7	0,4	326,4
2. Agriculture	31,1	80,4	-	-	-	-	-	111,5	9,2	52,3	0,6	0,7	0,5	-	165,8
3. Forestry	3,0	0,0	0,0	0,1	0,0	-	0,0	3,1	0,1	0,4	0,0	0,1	0,1	-	3,8
4. Building	2,0	0,1	0,0	1,2	0,3	0,1	0,0	3,7	-	-	0,4	40,8	-	-1,0	43,9
5. Transport and Communications	5,8	0,0	0,8	2,3	0,0	3,9	0,0	12,8	1,7	0,0	0,2	-	-	-	14,7
6. Trade	4,4	1,0	0,0	0,4	0,0	0,0	0,0	5,8	0,4	19,8	1,3	-	-	-	27,3
7. Remaining material output	1,4	-	-	-	-	-	-	1,4	0,0	2,2	0,2	-	-	-	3,8
Total	173,5	90,5	1,2	19,4	4,4	7,2	0,4	296,6	28,3	180,7	11,1	56,3	13,3	-0,6	585,7
8. Imports	16,9	1,0	0,0	0,2	0,6	0,1	0,0	18,7	0,1	5,1	0,6	5,3	0,8	0,5	31,0
9. Depreciation	11,6	5,2	0,0	0,7	1,9	0,3	0,2	19,9	-	-	-	-	-	-	19,9
10. Net output	125,1	70,6	2,6	23,8	7,9	19,8	3,2	252,9	-	-	-	-	-	-	252,9
11. Sums not balanced	-1,6	-1,5	-	-0,2	-0,1	-	-	-2,4	-	-	-	0,7	-	-	-1,7
Gross output	326,4	165,8	3,8	43,9	14,7	27,3	3,8	585,7	28,4	185,7	11,7	62,3	14,1	-0,1	887,8

Remarks : i) Figures 0,0 denote quantities less than 0,05 thousands of millions of Zlotys.

ii) Column 14 includes 0,2 under Industries and Handicrafts and the reminder is accounted for "Sums not balanced".

Please turn over

(NS)

GROUP B: TECHNO-COMMERCIAL STATISTICS (Special Paper III - Practical)

I . STATISTICAL QUALITY CONTROL (50)

II . OPERATIONS RESEARCH/ELEMENTS OF BOOK-KEEPING AND ACCOUNTANCY (30)

III . STATISTICAL METHODS IN BUSINESS (20)

Section 1. Statistical Quality Control (Special paper III)

(Answer question no.1 and any three from the rest of this section)

1. A large manufacturer of a detergent feels that the proper packaging of the detergent might increase sales. Therefore the organisation decides to experiment with 4 types of packaging. These represent the one factor of interest in the experiment. However, other factors that might be controlled in order to provide a more efficient experiment, include the location of the store (A, B, C & D) at which the detergent is sold and the day of the week (Monday, Tuesday, Wednesday & Thursday only).

As a basis for comparison of different experimental designs, let us suppose that packaging has no effect on sales. However, we shall assume that sales for different days of the week and in different stores do vary, and on an average they are as follows irrespective of packaging.

Expected Sale of Detergent
(Hundreds of Pounds)

Day of the Week	Store			
	A	B	C	D
Monday	16	17	27	18
Tuesday	19	29	22	17
Wednesday	31	26	28	28
Thursday	23	15	27	17

Consider and compare the following experimental designs in analysing the above data

(1) Completely randomized design
(2) Randomized block design
(3) Latin Square design.

(14)

2. A company samples 4 steel rods every half-hour and determines their average tensile strength. On this basis, it decides whether a process is or is not operating satisfactorily. Past experience has indicated that a desired process average is 600 pounds per square inch and the process standard deviation is 20 pounds per square inch. The company is willing to run a risk of .005 of stopping the process when it is operating satisfactorily. Specify the upper and lower bounds on the tensile strength, within which the process should be undisturbed.

(12)

3.(1) A company accepts or rejects lots of 10,000 machine parts on the basis of the following decision rule: Select a simple random sample of 200 parts. If 1% or more are defective, reject the lot; otherwise accept it. Plot an appropriate operating characteristic curve for this decision rule.

Please turn over

3.(ii) In contrast, suppose that the company modified its decision rule to require a simple random sample of 1000 parts, instead of 200 parts. Indicate approximately the effect of this change in sample size, upon the O.C. curve.

(8+4) = 12

4. (i) The number of accidents in a plant varies from week to week. The number of workers employed and the hours worked remain fairly constant from one week to the next. The firm wants to use a control chart to determine when and when not to take action on its safety programme.

For the first thirteen weeks the accident report gave the following:-

week ending	no. of accidents	week ending	no. of accidents	week ending	no. of accidents	week ending	no. of accidents
Jan 9	10	Feb 6	11	Mar 6	17	Apr 3	12
" 16	12	" 13	16	" 13	15		
" 23	9	" 20	17	" 20	16		
" 30	15	" 27	15	" 27	14		

Plot the control chart on the basis of this information.

(ii) Let us suppose that the following data are compiled for the ensuing 10 weeks. How does this affect the safety programme.

week ending	no. of accidents	week ending	no. of accidents	week ending	no. of accidents
Apr		May 1	12	May 29	7
10	1	8	10	June 5	18
17	10	15	8	12	4
24	27	22	16		

(6+6) = 12

5. The industrial engineer of a company must design a punch press operation, to stamp out inlays from sheet metal. He considers two alternative methods which differ only in the way the metal is oiled. Method B is more time-consuming than method A. However the engineer is uncertain about the proportion of defectives to be expected from each method. He therefore conduc a comparative experiment to provide a basis for decision. The engineer has 5000 inlays stamped by method A, resulting in 255 defectives and 4800 inlays stamped by method B, having 216 defectives.

Which method should he choose?

Since the method B is slower, the engineer feels that if the difference between the two population proportion is zero, he wishes only at 0.01 risk of adopting method B.

Please turn over (12)

Section II: OPERATIONS RESEARCH
(No candidate available)

Section II (Alternative): Elements of Book-Keeping & Accountancy (Special Paper III)

(Answer any two questions from this group)

1.(a) What is a Bill of Exchange and what are its advantages to a trader?

(b) Is agreement of Trial balance a conclusive proof of the accuracy of books of accounts?

(7+8) = 15

2.(a) What is a Bank Reconciliation Statement?

(b) From the following particulars, prepare a Bank Reconciliation Statement as at 30th June, 1968:

(i) Balance as per Pass Book on 30th June 1968 - Rs.1655/-

(ii) Cheques amounting to Rs.1623/- issued to suppliers on 29.6.1968, were presented for payment in July 1968.

(iii) The following cheques paid into the bank on 29.6.1968 were collected by the bank in July 1968:-

R.Bose	Rs. 360/-
A.Mitra	Rs. 150/-
R.Sen	Rs. 540/-

(iv) The following amounts were recorded in the Cash Book:

Commission debited by bank	Rs. 20/-
Interest credited by bank	Rs. 30/-

(c) The following figures relate to a business:

Bad and Doubtful Debt Reserve at 1st Jan'67	Rs.1400/-
Bad Debt written off during the year	Rs.1200/-
Maintenance of a Reserve of 5% on Sundry Debtors, which stood at Rs.18000/- on 31st December, 1967.	

Show the Reserve for Bad Debt Account in the Ledger.

(3+6+6) = 15

3.(a) Explain the following terms and indicate their treatment in the Accounts:-

(i) Fixed Asset, (ii) Wasting Assets

(iii) Intangible Asset, (iv) Secured Loans.

(b) Prepare a Balance Sheet showing at least five imaginary items on each side of such Balance Sheet.

(8+7) = 15

Please turn over

Section III: Statistical Methods in Business (Special Paper III)

(Answer any two questions from this section)

1. Suppose we have a random sample of 900 radio sets in operation in Calcutta during an important broadcast by the local station; 330 of these operating sets were tuned to the local station.

Obtain a 95°/. confidence interval estimate for the proportion of operating radio sets, which actually were tuned to the local station.

(10)

2. Suppose the financial manager of a company wishes to know the proportion of accounts receivable at each branch store, which are more than 30 days past due. There are 2000 accounts in all receivable at branch A. The accountant at braich A randomly selects 225 accounts and determines that 50 of them are at least 30 days past due.

What would be the symmetrical 90°/. confidence interval estimate of the true proportion of accounts receivable at branch A, which are at least 30 days over-due?

(10)

3.(a) A study of demand (d_t) for the past 12 years ($t = 1, 2, \ldots, 12$) has indicated the following:-

$$d_t = 100, \quad t = 1, 2, \ldots 5$$
$$= 20, \quad t = 6$$
$$= 100, \quad t = 7, 8, \ldots 12$$

Compute a 5-year moving average and discuss the outcome.

(b) A company desires to forecast sales for the coming year. It can use the forecast as a basis for decisions on the budget for the year, production schedules, sales quotas, advertising plans, procurement policy and labour needs. In the short-run forecast however, it is necessary to know the seasonal pattern of sales during the year, in addition to forecasting annual sales. As a first step, the company extrapolates the past trend of sales. This gives an estimate of trend. It is then necessary to modify the trend forecast for cyclical and possible random effects, and for possible changes in past trend forces in the future.

Outline seven or more modifying factors which should be considered at this stage. (eg. Government control).

(5+5)=10

Please turn over

GROUP C: BIOMETRIC METHODS
- Special Paper III

(No candidate available)

GROUP D : DESIGN & ANALYSIS OF EXPERIMENTS
- Special Paper III

(No candidate available)

GROUP E : SAMPLE SURVEYS (Special Paper III - Practical)

(Answer all questions from this group)

1. A sample of ten villages was drawn from a tehsil with probability proportional to size and with replacement - size being the 1961 census population - for estimating the total cultivated area in that tehsil.

Table : 1961 census population (x) and cultivation area (y) in acres for ten sample villages

village	x (units)	y (acres)	village	x (units)	y (acres)
(1)	(2)	(3)	(1)	(2)	(3)
1	5511	4824	6	7357	5506
2	865	924	7	5131	4051
3	2535	1948	8	1654	1080
4	3523	3013	9	1146	830
5	8308	7678	10	1165	1013

Using the data given in Table above,

i) estimate the total cultivated area and its relative standard error (r.s.e.); and

ii) obtain the sample size required to ensure an r.s.e. of 2%. (17+15+17)=35

Please turn over

EITHER

2.(a) For a Socio-economic survey, all the villages in a region (including the uninhabited ones) were grouped into four strata on the basis of their altitude above sea-level and population density. From each such stratum, five villages were selected with equal probability and with replacement. The data on the number of households in each of the sample villages are given in the Table below.

Table: Number of households for 20 sample villages.

stratum sr. no.	total no. of villages	total number of households in sample villages				
		1	2	3	4	5
1	1411	43	98	10	0	15
2	4705	50	62	84	170	56
3	2558	228	110	139	334	62
4	14997	17	25	36	25	15

(i) Estimate the total number of households and its relative standard error.

(ii) Estimate the gain, if any, due to stratification.

(iii) Compare the efficiency of the present allocation with that of proportional allocation, keeping the total sample size fixed.

(5+10+10+10) = 35

OR

2.(b) For estimating the total (y) of current population in a region, two sub-samples, of six villages each, are selected circular systematically from each stratum, with independent random starts. Using the data given in Table below, obtain a ratio estimate of y, taking the previous census population (x) as the supplementary variable and estimate its relative standard error.

Also compare its efficiency with that of the usual unbiased estimate.

Table: Total number of villages (N) and sample totals of x & y

stratum number	no. of villages N	sub-sample 1		sub-sample 2	
		x	y	x	y
1	2044	3722	3935	3456	3541
2	1304	3625	4033	4171	4549
3	1265	2769	3050	3746	4048
4	1252	3180	3498	4323	4778
5	4264	3522	3819	3314	3636
6	1598	2827	2936	3550	3652
7	810	8603	9596	7285	7965
8	567	8323	9135	9595	10014
9	500	9019	9772	11073	12112
10	486	7404	8185	6981	7610

(10+15+10) = 35

EITHER

3.(a) The table below, giving the percentage distribution of estimated number of persons by monthly per capita expenditure classes and by population zones, is a modified version of a table based on the data collected in one of the rounds of the National Sample Survey.

Table: Percentage distribution of estimated number of persons by monthly per capita expenditure classes and by population Zones.

per capita expenditure class (Rs.)	population zones*						all-India
	1	2	3	4	5	6	
0-8	26.44	17.40	30.18	27.00	30.61	22.78	52.03
8-11	27.95	16.10	18.25	22.32	25.20	17.88	20.62
11-13	9.18	14.44	10.80	9.92	12.46	7.14	11.33
13-15	6.88	9.35	7.83	9.13	9.82	8.57	8.57
15-18	7.48	10.87	9.23	9.63	6.80	12.03	9.34
18-21	6.03	6.58	6.66	7.80	6.34	7.93	6.71
21-24	5.63	5.68	4.47	4.63	2.37	7.30	4.99
24-28	2.35	4.50	3.91	4.56	2.26	4.36	3.66
28-34	3.48	6.77	4.94	2.80	1.12	4.95	4.44
34-43	2.47	4.19	2.31	1.48	8.6	3.43	2.68
43-55	1.18	1.11	0.95	0.73	0.71	1.04	0.99
55-above	1.29	3.01	0.46	-	1.45	2.59	1.63
all classes	100.00	100.00	100.00	100.00	100.00	100.00	100.00
no. of sample villages	102	166	112	84	122	120	706
no. of sample households	276	444	316	202	328	292	1869

(* 1 - North India; 2 - East India; 3 - South India; 4 - West India; 5 - Central India; 6 - North-West India).

Carefully examine the table for consistency of the figures and also give your observations on the results from the point of view of regional similarities and differences.

(30)

OR

3.(b) In the 1971 Indian Population Census, it is proposed to canvass the following items for all ever-married women living in a 10% sample of area units, (villages/urban blocks) for studying fertility patterns:

1. age at marriage
2. husband's age at marriage
3. number of children - residing with her
4. number of children - residing elsewhere
5. number of children - born alive but now dead
6. total of (3), (4) and (5)
7. number of children born dead
8. live birth(s) in the last one year
9. still birth(s) in the last one year
10. date(s) of birth for (8)

If the tabulation of the census fertility data is to be mechanised, give an idea of the work-load involved at the processing stage, in terms of the requirements for scrutiny, editing, coding, punching, verification, tabulation and post-machine scrutiny.

(30)

Please turn over

(over)

GROUP F: TECHNIQUES OF COMPUTATION (Special Paper III-Practi

(Answer all questions from this group)

1. In order to study the employment situation in the count a sample survey was conducted in Urban India. Each State was first divided into a number of strata and from each stratum a few urban blocks were sampled. Within each such sample block, certain number of households were selected and data on employ particulars were collected by interviews.

The following gives the details collected from one person in a sample household:

Household Identification:

State code	...	23
Region code	...	1
Stratum number	...	5
Sample Block number	...	3
Sample Household number	...	2

General Information:

Household religion	...	3
Household size	...	11
Total Monthly Household Expenditure (average)	...	1531.72

Employment Particulars:

Person serial number	...	12
No. of days employed (last month)	...	27
Earnings	...	192.47
Occupation code	...	173

For processing these information, two types of cards are made out for each household. (i) One household master card contain household identification particulars and general information particulars. (ii) Individual cards, one for each person giving the household identification particulars, person serial numbe number of days employed, earnings and also occupation code.

The card designs are as follows:

Household master card:

item	card columns	remarks
CDI	1-4	Punch X X X
State code	5-6	
Region code	7	
Stratum number	8	
Sample block number	9-10	
Sample household	11	
Household religion	12	
Household size	13-14	
Total monthly household Expenditure	15-20	Rs. P.

Please turn over

Person Card:

item	card columns	remarks
CDI	1-4	Punch 1 2 3 4
State code	5-6	
Region code	7	
Stratum number	8	
Sample block number	9-10	
Sample household	11	
Person serial number	12-13	
Number of days employed (last month)	14-15	
Earnings	16-21	Rs. P.
Occupation code	22-24	

Punch one master card and one person card for which the data are given above.

(15)

2. A deck of master cards and a deck of 'person' cards described in Q.1 are supplied

i) Arrange the master cards in ascending order of household identification fields (state X region X stratum X sample block no. X sample household)

ii) Arrange the 'person' cards in ascending order of household identification X person no.

iii) Merge the two decks together so that the resultant deck is in ascending order of household identification with selection of unmatched cards from both decks. Within each household the 'person' cards should follow the master card.

iv) Prepare a statement showing for each State, total number of households, householdsize and total number of days employed (last month).

(7+8+20+20) = 55

EITHER

3.(a) In a manufacturing organisation weekly gross pay (P) for a worker is calculated according to the formula:

$$P = \Sigma h_i \times r + \Sigma m_i \times h_i$$

where h_i = hours worked on the i-th job

m_i = hourly incentive rate for the i-th job

r = hourly rate of the worker, including D.A.

P = total pay for the worker

summation extends over the total pay period (say one week).

Net pay is computed after making the following deductions

i) P.F. contribution (10% of P)
ii) E.S.I. scheme (5% of P)
iii) Income Tax (4% of P)

The staff file contains all information e.g. roll number, name, hourly rate of the worker. The incentive rate file contains informations about job no., hourly incentive rate.

Please turn over

Assuming that there are about 1000 employees and 200 types of jobs, describe a suitable scheme for preparing the pay bill cum acquittance roll, using punched card equipment.

Give proposals for card designs, in flow charts and time estimates, naming the types and makes of machines you propose to use. (Control panel wiring charts are not required to be furnished).

(30)

OR

3.(b) You are supplied with a deck of cards which contains informations collected from primary schools in India. The items of information contained in the cards are as follows. (one card per school):

Sl.no.	Description		Card column	Remark
1	CDI		1-4	S S S S
2	State		5-6	
3	Town		7-8	
4	School number		9-10	
5	Number of students:			
	i) Class I	(a)	11-12	
	ii) Class II	(b)	13-14	
	iii) Class III	(c)	15-16	
	iv) Class IV	(d)	17-18	
6	School fee	(e)	19-21	Rs. P.

Using a calculating punch 602A or any other make of your choice,

i) Calculate the total number of students (T) for each school (where T = a + b + c + d) and punch the information on column 22-24 of the same card.

ii) Multiply total number of students (T) by school fee (e) to get earnings (f) and punch on columns 25-30 (Rs. P).

iii) Sort the cards: State X Town X School

iv) Prepare a statement showing the total number of students and total earnings for each state and town.

(30)

GROUP G: STATISTICAL INFERENCE (Special Paper III - Practical)

(Answer any four questions from this group).

1.(a) Let X be a binomial random variable with $P\{X = 1\} = \theta$ and $P\{X = 0\} = 1 - \theta$. You are to test H_1 ($\theta = 0.4$) against H_2 ($\theta = 0.6$), on the basis of 10 independent observations on X. The loss matrix is given below:

States of nature	Decision	
	H_1	H_2
H_1	0	25
H_2	25	0

Find the minimax procedure and compute the least favourable prior.

Please turn over

1.(b) Let $X = (X_1, \ldots, X_{100})$ be 100 independent observations on a random variable which is normally distributed with variance $\sigma^2 = 25$ and mean θ. Let H_1 be the hypothesis that $\theta = 0$ and H_2 the hypothesis that $\theta = 2$. Let $p = (p_1, 1-p_1)$ where p_1 is the a priori probability for H_1. Plot the Bayes risk as a function of p_1 if the loss-matrix is given by

States of nature	Decision	
	H_1	H_2
H_1	0	25
H_2	10	0

Hence obtain the minimax procedure approximately.

(12+13) = 25

2. Let $X_1, \ldots, X_n$ be independent and identically distributed with common density $f_\theta(x) = \frac{1}{\pi} \frac{1}{1 + (x-\theta)^2}$. Let T_{1n} and T_{2n} be the sample lower and upper quartiles. Using large sample approximations, find the asymptotically best unbiased linear estimator of form $b_1 T_{1n} + b_2 T_{2n} + b_0$, for θ. Compare its efficiency with that of the sample median and the maximum likelihood estimator for θ.

(25)

3.(a) Let $X_1, \ldots, X_n$ be independent and normal with variance $\sigma^2 = 1$ and mean θ. Let H_0 be the hypothesis $\theta = 0$ and H_1 that $\theta = +1$. Find the smallest n for which the Uniformly Most Powerful (UMP) unbiased test of H_0 of size $\alpha = .01$ has error second kind $\leq .01$, under H_1.

(b) $X_1, \ldots, X_n$ are as in 3(a) above.

Consider each of the following sampling schemes:

i) Take a sample of size 9
ii) Take a sample of size 2 or 16, with probability $\frac{1}{2}$ each.
iii) Take a sample of size 0 or 18, with probability $\frac{1}{2}$ each.

For each of the sampling schemes described above, find the most powerful test of H_0 ($\theta = 0$) against H_1 ($\theta = 1/2$) of conditional size α_i corresponding to the i-th sample size: for (i) and (ii) take $\alpha_1 = \alpha_2 = .005$, for (iii) take $\alpha_1 = 0$, $\alpha_2 = .01$. Calculate the power at $\theta = 1/2$, in each case.

(10+15) = 25

4.(a) Y is a 4x1 random vector normally distributed with mean vector $(\mu_1, \mu_2, \mu_3, \mu_4)$ and dispersion matrix $V = u^{-1}$ where

$$u = \begin{bmatrix} 2 & 0 & 1 & 0 \\ 0 & 5 & 0 & -3 \\ 1 & 0 & 1 & 0 \\ 0 & -3 & 0 & 2 \end{bmatrix}$$

On the basis of 10 observations with sample mean vector (1.1, 0.5, -2.1, 3.2),

i) test H_0 ($\mu_1 = \mu_2 = 0$) at level .05
ii) also test H_0 ($\mu_1 = \mu_2$) at level .05

Please turn over

4.(b) Given the data:

y	12.1	11.9	10.2	8.0	7.7	5.3	7.9	7.8
x_1	0	1	2	3	4	5	6	7
x_2	7	4	44	6	4	2	1	1

fit the model

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$ where ϵ is normal with zero mean and variance σ^2 and x_1, x_2 are non-stochastic. Find a 95% confidence interval for σ^2.

(12•13) -

5. Let $X_1, \ldots, X_n, \ldots$ be independent and normal with variance $\sigma^2 = 1$ and mean θ.

(a) Construct the sequential probability ratio test of H_1 ($\theta = 1$) against H_2 ($\theta = 2$), with $\alpha = .01$, $\beta = .05$. Calculate the ASN at $\theta = 0$.

(b) Let $Z_i = 1$ if $X_i > 0$

$= 0$ if $X_i \leq 0$

Construct the sequential probability ratio test of H_1 ($\theta = 1$) versus H_2 ($\theta = 2$), based on $Z_1, Z_2, \ldots, Z_n, \ldots$ for $\alpha = .01$, $\beta = .05$ and calculate the ASN at $\theta = 0$.

(12•13) - 2

GROUP H : PROBABILITY THEORY
- Special Paper III
(No candidate available)